

Fig. 2d



Fig. 2h

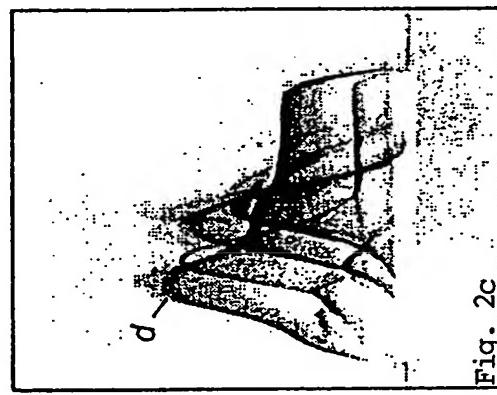


Fig. 2c

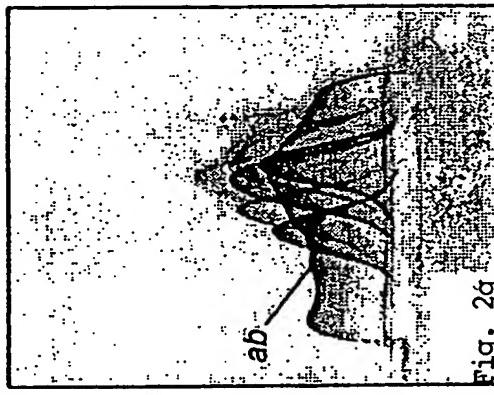


Fig. 2g

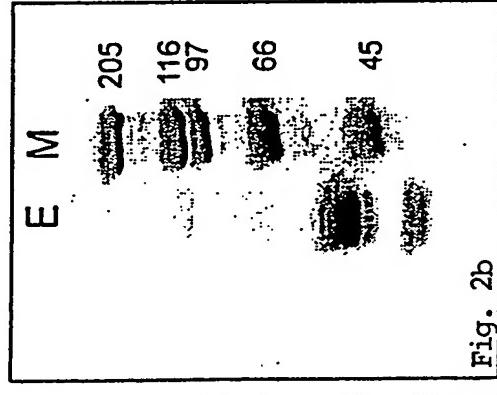


Fig. 2b

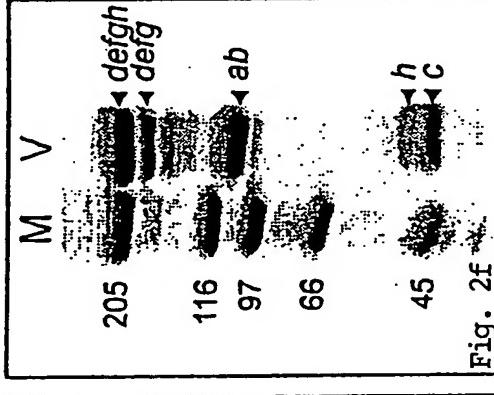


Fig. 2f

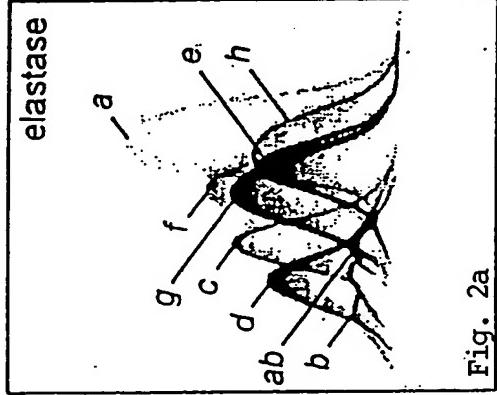


Fig. 2a

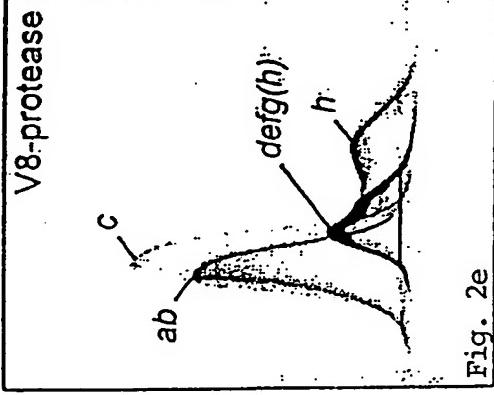


Fig. 2e

Nucleic Acid Molecule Comprising A Nucleic Acid Sequence Which
Codes For A Haemocyanin And Comprising At Least One Intron
Sequence (Serial No. 10/049,988; Inventor: Markl Jurgen)

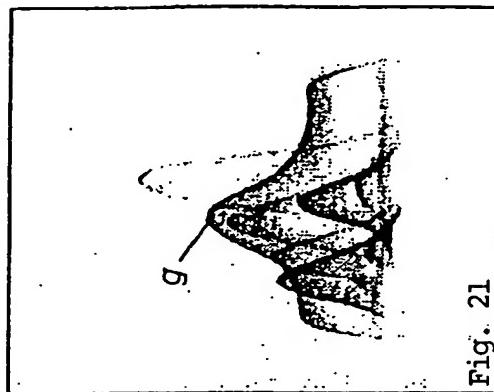


Fig. 21

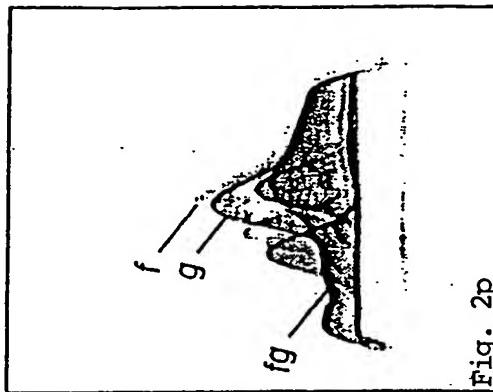


Fig. 2p

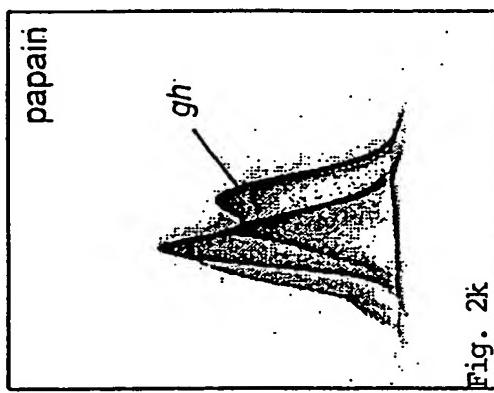


Fig. 2k

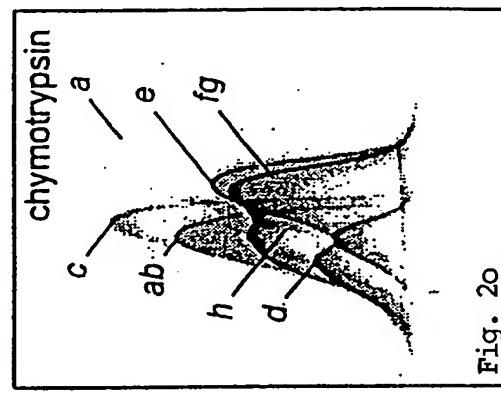


Fig. 2o

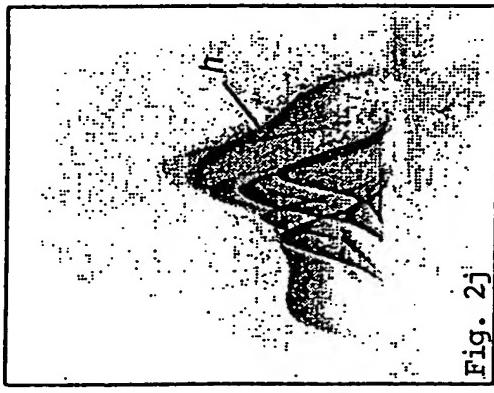


Fig. 2j

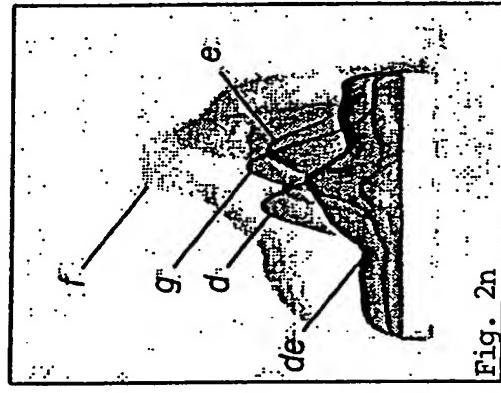


Fig. 2n

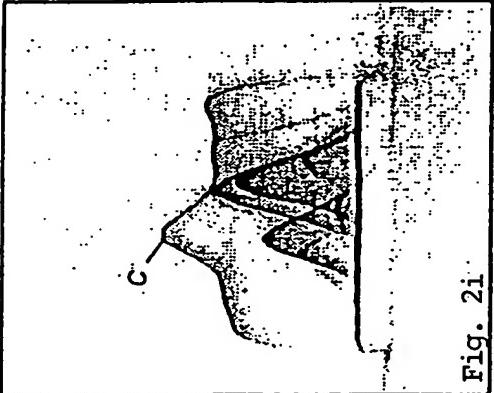


Fig. 2i

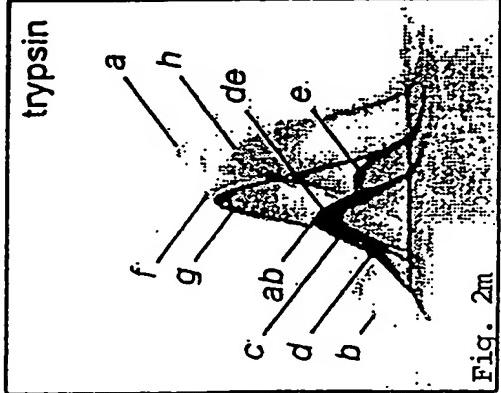
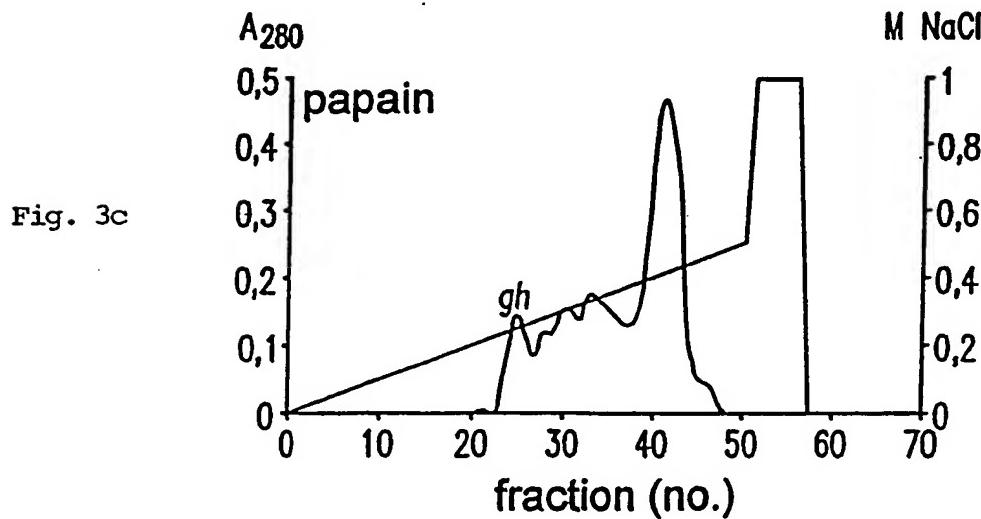
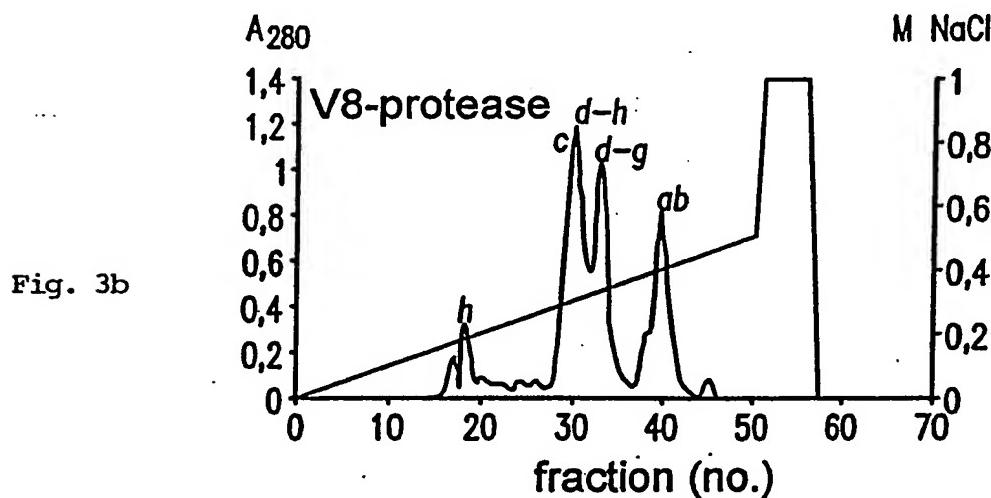
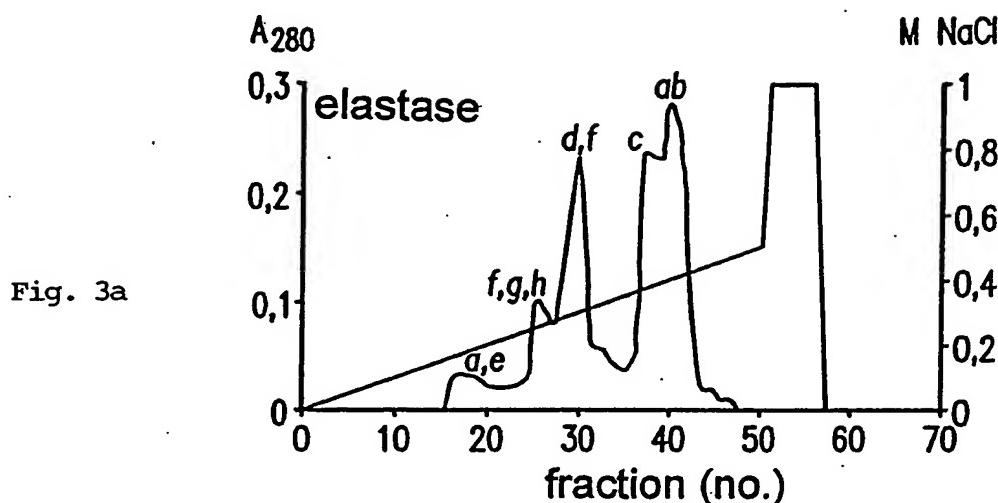


Fig. 2m



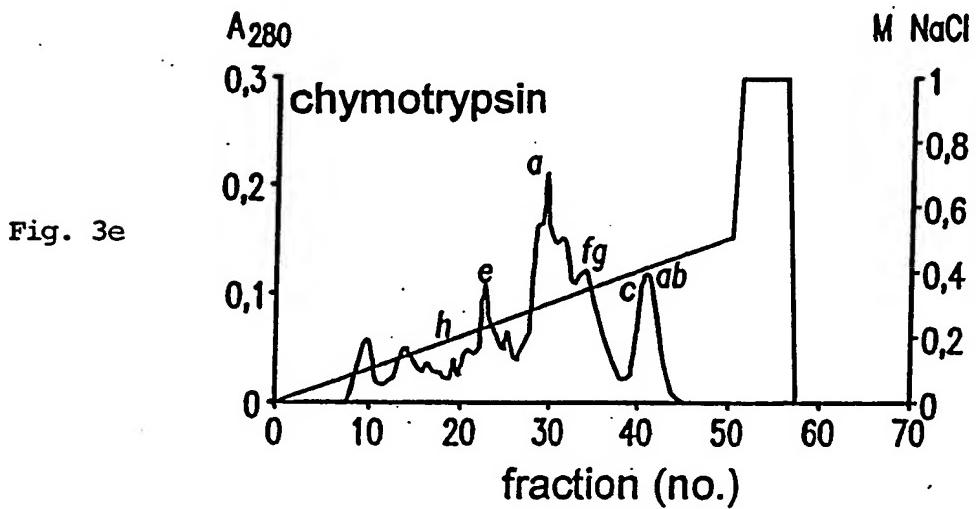
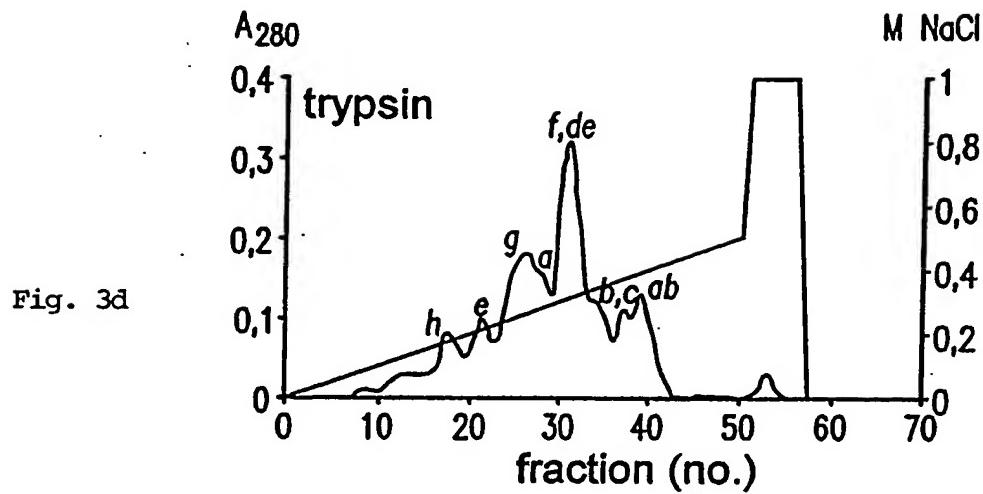


Fig. 4aGenomic sequence of the HtH1 gene

SIGNAL PEPTIDE SEQUENCE 1S-1 (1st part)

GGCTTGTTCAGTTCTACTCGCGCCCTTGTG

INTRON 1S-1/1S-2 (SEQ ID NO:109)

GTAAGTCAACGTCTTGTAAAGTTGATGCATATCTATCATTGCGTTAAAATACCA
 TTACAACCAACGTGTCTCTATTGGTCTCACCTGTTAACGTATATATTGTTTAAATGT
 GAAAATCTGAGATTATTTCATTCCGTCAATATTGTAACAAATACTATACAAATAAAATT
 GCTTCAGCCTATTGCATTGGCAGTTTCGAGAATAACGAGGGAGGCGTACATAAAATA
 TAAACCAGTGTATATTCAAGCATGTTATAATTCTTATAGATTATAACATCATATCAA
 AACACCAATCTGGATTAAACCCGTGAATCCAAGTATACCAATTACGGAACCTTATCA
 TGTTTATCAAAGGTTTAGATGAGGGTAAAGAAGTCCGAGCTATATTGCGATATCAG
 CAAAGCCTCTATCACGTCTGCACACAGGGCTGGTATCTAAACTCGAATCCACAGGAAT
 AAATATTCAGCCGATAGAGAACAGTCGGTGGCTATCTGGTACAAAACAAGTCCAAA
 ATCTGCATTAGCCGGTGTCCCCAAGGCTCTGTCTGGGCCACTATTATTCACACCTA
 TATAAACGATTCAACTAATGGAATATAAGCAACGTAAACCTCACCGCAGATGAAACACT
 AAGTTATAGACAATCCGTTAAAACCCAGCCACTGCTTAATAATGACTTAGGCCGTCTT
 CAGACTGGGCTAGTAAGCGGCAGGTTAAATTCTACCTTGAAAAGACAGAACCATGGTAT
 ATTTCAAAAACACGAATGCAAGTCCTAAACTCAACTACTACTTGATGATACTGGGATT
 CTAAAGTGTGTGAACAAAAACACATTGGCCTGATCCTACAAGATAACCAGACAGAACCA
 TGTTTTTTCAATAACACGAATGCAAGTCCTAAACTCAACTACTACTTGATGATACTG
 GGATTCTAAAGTGTGTGGTGAACACAAACACCTTGGCCTGATTCTGCAAGATAATGGAA
 AATGTCAGAAACATAAGCAAGTTGATGTTGGGTTCTGGGGTTGTGACAACACCGAAA
 GACCCTGCAACTATGTTAGCTCAAAGGGTTTACACCCGGTACAAGTGGGATCGACC
 CAGGCACCTTGCCTTGACAGCTCGCTTCAAAAATCTCAATTGAAACGAAATC
 TAATAATTGAGCGATACAACCGTTTCTATAATGCTGTGGTACCGCATACTGTGGA
 AACATCTGTCTACCCATTGGTAGTCCCCATAAAATGTATTTATGTTATAAACACAAT
 GTTTATAGGGTTACAGTTAGAAGAACGATTCTGCTTATGGCTAATGTCATTGCTT
 ACTATTGTGCAAAGGCATATTACAGGTCTTTAGGAAATTAAATACTGTTAAATCACAT
 ACACTACCGGTAACTCTATTATGCTTATCCTGCCAACATTCTGCCAAGCAAACGCATGA
 AAGTTAAAGCTGAGTGTAAATACTGATTGCTGTGTTACTTCACAACCAGTGGACTGAAT
 ACAACCATGTTCTTGAAAGTCACAAACATCCAGTCGGTTCTAATGTTAAAGTT
 CTAGTTCTAAAGAGCATGACGTAATGGTAATGGAGTTCAATGTTCTATCTAAT
 GACTCCTAGTCGTTACTTTTAATAAAACATCCATGTGTTAATGTTGGCCACAGAT
 ATAACAAGAAAGAAATCGGATAAAATCACATTGACCAATCGAAGGCTGCCCTCC
 CTAATCCTAATCATTGTCGCTCAAAACATACTCAACCAGACATTGAACTATGTATA
 TATCAGAATGAAATGGTAACAATAAACTGTATGTTGACAGACAGAATTAGGGTGAATC
 TGAATACCAACTATTGTCACATATGAATATGGATAAGCTCTGCGCGTGCCTGGCGGT
 GTAGTGCCTGTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
 TGTGTGCGCGTGTGTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
 TGCACAGACATGTGGTTGAGACACACTTGATTGTCAGTCAGTCAGGATTATGTCCTCAACCGAG
 TGTAGTCCTTAAGTGTGCTGGAAACAAAAACTGCGTTGGGTTGCATGCCCTCTGTAGC
 AAGCTTGGACCGCGTCACGCAGCTCTGATACCACTGATTGGCACCAGTTCATCGGTCTC
 ACGCGAATATTATGCTATGTTGGCGTATCATAACCATAGGTTGGAACGTTCAATACTG
 TACCGAGCTGGCGTGTGTCACAAAGCTATGATAAGATGACAACACGTCTGGCATCTTGT
 TTCCTCGGTATCACCGCTGTTATGCTATGTTGGCTATCACACCTTAGGTTGGAAAGT

Fig. 4b

TTCCACATTTCCAGCCTCGTACATGTTCCCTTGTTCCTAGTTATCAGCATAC
CGTATATTCTATTTAATGAGCATTGTATTTCTACAG

SIGNAL PEPTIDE SEQUENCE 1S-2 (2nd part)

GTGGGGCTGGAGCAG

INTRON 1S-2/1A-1 (SEQ ID NO:110)

GTGAGTTCTAACATTGTACATGGTACATGGATATACGCTCAGTGGAAAGCAGGATATC
CCCTGGTTCAAGTATTCACTTGTACGCCAAGTGTGATTCCCAACATGGAATACTGT
CATATAGTAAATTGATACACTACTTACATTTAATTCTCCACTAAACGTCAACGTCCTTTA
CTTCATGGCCACATGGTCCGTATTAGTGAGTGAGTGAGTCAGGGCATAAGTATTTAACG
TCAAATCAGCAATATTCCAGCCATATTGTGACAAGAATTGAATATAAATAATTACTTA
TAATGCTTATAAAATATAAATTATATAAATACCTATAACTATAAATTAGTTACTAGTAT
TTATCAAAACATATTGCCACGACACTGCACGCCGATACTCAAGTGTCTCACCTCAAG
CGTGTAACTCCTCATACTCTGTAAATAAGTATGTACACTAAGTGAGTGCTATCATCTCCAT
GCTTCATTAGTTGTCAGATGCGTGTATCCATACGAGTACATTAGATTATGGGATCCA
GAGCTTCTTATCTCAAGTATTCCGATTGTAAAGCCATACTACTTCCCCAATGACTGAC
GAGACAGATGCAACCGTTCTTCCTGACTAGGTGAGTGCCACTGATAAAATCATTAT
GCCTTAACATTAGGAATGTTAGCAGTGCACATGTTCAGAATTGCGACCTATGGTTGT
AAAGATTACAAACTTACAACTTACTTGAGACAGGTTCCATATGCGTATCTGAAATAGT
GTGAAGGTATCTGATTGATGCAATACAGACATATAAACATATTGCGCCCTGCTATT
CCGGAAAGGTCTTTGTATGTAACGTTCTTAATGGACACAAACGGAATTATTAGTTAA
ACATACTCAACAAACTATGTTATTGCAATGGTAGCACCAGACAGTGG
TTCGTAAAAGTAGAACATTCTGACATAAGAAAAATCATTGGCTTAAATATGCAAGT
TACTTGTCTCAACAAACAGTTTATACACATTTCAGAGAACGGGAATCCGCGATGACA
ATATCAACGAGTATACAGAAATATAATTAAAAACGATGAGTGCTGGCAAGGGAAAG
AGCGAGATTGCAAACAGGGGGTGGTGTGAGCTGATCGTGGAGAAACGTAGATTG
AAAGACAAGATGACATCTAATGATCCAAAATCAAACACAGGATTAACTGGGATGCAAGA
GAATGAATATCTCAAGCATACTGCAACACTTCATGAATGCATCTCAAACATTTCGTCA
GATCGGATGCGATGAAGATTGTAAGCAATGGTTAAATTGTCCTAAACGTTAGTTGG
AGATGTATGAGGCTAGGCTGTATGTTGAAACGAAACCATTAAACATTGTTGTCATGATTA
TTAATATTTTCTTTATAGATGTACAATAAAATTGAAACTAAACATTCCCTTTA
TTGTTTGTATTCACCTGTTCATGGGTATGTTGAAAGATCGTGAATTAGTTGGCAT
TCACAAGTTGAAAAAGGTCACTCAGTTGATTCAAGTTATGTAACCTCTTATCTGA
CGCTCCAAAATATGTATAGCCTGTCATCTGTCGGTATGTTGATATTCCCTACTTCAGGG
TAGGGTAGCATTAACTTACAAAACATAACGTTACAGTGTACCAAGATTCAGTCACCTCAGAGAT
GATAATGCATGCGATATGATAGGTCAAACATTGATCAATCACAATGAACCTATGG
ACCCTGAATCGGAATGATACGTTACACTTAAAGCAATTCAACAAATATGACTGTCACCC
TTTCAGGTAAATGTTGACGGACTACGATAGTGCTGAACAGCAGGAGAGGCAACATGG
TTCGATTGAGACAGGTTAGTGTATTGTTGCGAATTAAAGGTTCTGAATCACAATA
GACACGGTTCAAGTTAATGGATAAAACCAATCATTAGATAGATAGAGATTAGTCGCGATATT
GCTGGATAAAAGCTTAGTGGACGTTAAGTCCCCTCAATCTCTCATTTCCTCAA
ACAGTTTAATTCAAGGCTCATGACAAGGTCGTACTGTTGCAAAGGATTCTACTTCAAGCA
GAGATGTCTCATGAATACAGTACAGGGTTTGAAGTTATCCAGTGCAGCGCTGGCACC
ATCTCTGCATGCGAATTATACCATGCCGCTCTAGGCTATTGTTATTAGTCTGTAG
AATTAAATTGCGAGTTGCAAATACTGCTCACCATTATGCGCTCAACCCAGTTGGTA
CATGCGATTACACAATTATGTTATTGTTATTGAGAAATCAACCTGAGATGTTGAATCGGGAGCT
GCGCTTATTCAATGGTGGACTCGGAAGGGAAGTAACCGCTGATGAGGCAAAACAATAACG

Fig. 4c

CAAACATATGGAAGTGGAACTCTTGAACCAGTATTATGTTGTGAGCATGTATGTGT
TAATTGACCATTGAAACAACCTTACTATTCTATTCTATAATGTGTTAGATTACATTG
AATTAAAAGAGATGAGTTAAGATATTAATATTTCCTTATAGTCTGCGTGATTGTA
GGGCAATATTATGTATGTTGTCATTTTCAATTTCATTTGAAAGGTATATCATAA
GATTATTATTATCATTCTGAAGTAATGTATACATATATATATGTCTGAGTAGCTTATT
TTCAATTATTATCATCCGTATCCAATTTCACGAAAGTATAAGAAATAACGAGA
GAGAGAGAGAGAGAGAGAAAAGACAGAAATGAAGTTAGGAGATATNAGTTATCAAGAA
AACACAGTTGAATTTTGTTAGACAAGATATCATATCAATAACCTCGCACTATTAC
GGGAATAGGCCGGCGTTCCATATGCACAATGAATCGTCAAGTAAACATTAACATT
AAAATACTCCATATTAAAGTTGATCTACCTCTTGTATTATTGTAGACTATTAGACAG
AAGTCGACAGTGACACCAGCAACCAGATATCATACCCAGACTAAAAAGCTGTTCTTG
ATGTTCAATTATTCCATTCCATTATTCCCTTATTGGTTCCATTATCAAACATT
ACCATCTGCACCACTGGGAGATTGATATGTTGTATTATTATTTCTTGTACTACAAT
ATCAAGAATGTATAGGAGCTATTCTGTTCTAAAACCGGATAGATCCATAATTCCAT
TTTGGGATAAATGAAACTAAACACAATTACAGTAAACACAGGTGAGCAAGTTGAGT
TTACGCCGTTTTAGTAGTATTCCAGCAATATCGGGCGGGGACACCAGAAATGGGCT
TCACACAGTGAATGCATGTGGGATTGCAACCCGGTCTCGGCGTGACGAGTGAACGCT
TTAGCCACTAGGCTACCCCACCGCTATTATAGTTAAGACGAATACTTTCTCAAGCCT
CAAATATGTCCATTCTAGAGAGACTGAATCTGATCCTGAATCTGCGGACCGGTCTGAAT
ATCATCCCACAACTCATTGACAAAGTACCTGTAGATTGTCAGTTCAAAGACAGATTTC
ACAACCCATTATATTTCCTGCTCATTAAGATATTCAAGACTCACTCAAACGTCTAAA
TGATTAACTCTACTTGAGATGTTAACCTTATTGATGCAATTTCGTTCTGAAT
TCCTGTATAAAGTAAAGCAGGTAAACCTAACCTAACCTGTGATTATTCATAGTTTG
CGATCAGATTGAAACCGGAATGCACAGTGAAGTGTGGCATACATCTTCACAGAGATAC
TGGATACTAGGTGGTACAACCGCATTGGCTTGTGAAAGGATATTAGTGTGTTATGAGAC
TGACTCATGTTCAATGCTTAGAGCGGAATGATCTGGTCTTCATGAAAATATTGTGTT
GAAGTAACCCCCCAGTCCCTAACAGAACGTGGGAAAGCAGATGGATATGCCAAGACATC
TTCGCATGGTGTGAAGATGATCGTTACACATCTGAGAAAAAGTTATTCTGTGAAGAA
TATGCCAAAGCATCACTGTGAGTGTGTTGAAGATGTGATATGGCAACACCGCAGCGTGTAA
TTATGCTTGTGTATTCTGAAGATCCGTATGAGCATGGGCCAAACTATCAGTTAAA
TGGCTATGCGAAGATCTCCGAGATGGTAAACACATATTGGCCATTTCCTTGTAAG
TGGCGACACAGAAGATCCCCCTGATTGTGTGGATGAGGACACAAAAACGGTCCCCCT
CCTTGCTGATGCTAATGACGCCCTGGAAACATGAAAGACTTCTCTCCAGCAAGCAAG
CCACCAACCACAAGTTGATAAAATGCTTGGACTTGAATACGGCAGTTGGACAGATAC
ATACAGCCAGAGAGGGCGAACTAAACATCTAAACATGGAAAAGTGTAAAGACAGGCT
TTGTTGTACGACGTACGTAAATTCTGAATGTTGAAAAGGTAGAAAATTATTAATCT
TTGAAACCTCGCTCTGTTGTTATTGTCACATGGTATCCAAAAAA
GGGCAGACACATTGTTTAATCTTAGCCAGGTCAATTAGCCTGCGCCAGACTCAT
TGTATCTGGTGAAGGCTATAGGTGGCACGTCTCTAAGATGCTATGCTATTCTTACCAAG
AATCCAATGTAAGAGTTCAAACGCATGGTCTGTTGATTGTGATTCTTCTTAGCACC
TCTCTCCTACCCAGAGTTCACCTGCACTGCTCTGACTCACAATAAGCTGACGTGCTGTC
ATATATGTGCAACATTGTATACGTTGGCGTTAAGCCAACTCACTCCGCTGTCTTTGG
CAG

DOMAIN 1A-1 (1st part of domain a)

ACAACGTCGTCAAGAAAGGACGTGAGTCACCTCACAGTTGACGAGGTGCAAGCTCTCAGG
GCGCCCTCCATGACGTCACTGCATCTACAGGGCCTCTGAGTTGAAAGACATAACATCTT
ACCATGCCGACCGAGCGTGTGACTACAAGGGACGGAAGATGCCCTGCTGTGCTCAGG
GTATGCCAGTTCCCTCTGGCACAGGGCATATGTCGTCAGGCCAGCGGGACTGT

Fig. 4d

TGTCCAAACGGAAGACTGTCGGAATGCCTTACTGGGACTGGACGCCAACGCTGACTCACT
TACCATCTCTGTGACTGAACCCATCTACATTGACAGTAAAGGTGGAAAG

INTRON 1A-1/1A-2 (SEQ ID NO:111)

GTAACTACAAACGTCGCCATTACATACAGGAGAAAATATACAATTGTGTTGTAAGAGCGG
TATACTGTTGCCAACTGTGTAATTGAAACGTTGATGATGGTGTCTTGATTTCAATT
GTATGCACTTAGACATGATCAATGTTCTGATGTTGCAAGGATGTTGGTGTGTCAC
CAAAGATCAAATTACATATGACGTACACAGAGCAAGAACCAACAGTAAGAAGTCTGTATG
ACTTCGCTCTAAAAGCAATGGAAAAATATTTCACTTAACACCTAGCCCATAATCACGC
ATATTAGATTATTCAAGCGATGTCAACATGTTTAATATCAATCTCATGGTCTGATAT
TACCGGAGACATGCAACAGGCTGCCATTAGCCAGGAATCTTATGAATATGTGCATAT
TTTTCTTGATTCTGTATGACGAGAAATATTGGGAGGCAAAGATTGTGTTTCAGAAC
GAATCAGGGTATCAGTGACATCGTCACTGCATGGCTACAATATTGCTGATGTGACTGTT
CTCCAAGGATTTCATCTCACTGTCTGACTTTGAATCTACAAATTGTTAAAGTTAT
GACAATTACCCCTGCCTATTGTAAACGAAATATAACATGAGTGTGTTATGCTGACAG

DOMAIN 1A-2 (2nd part of domain a)

GCTCAAACCAACTACTGGTACCGCGGGGAGATAGCGTTCATCAATAAGAAGACTGCGCGA
GCTGTAGATGATCGCCTATTGAGAAGGTGGAGCCTGGTCACTACACACATCTTATGGAG
ACTGTCCTCGACGCTCTCGAACAGGACGAATTCTGAAATTGAAATCCAGTTGAGTTG
GCTCATATACTGCTATCCATTACTGGTTGGCGGTAAATTGAA

INTRON 1A-2/1A-3 (SEQ ID NO:112)

GTAAGTTGGTTACAGTTTCAATTATAAAACATAGCAGTTAAGTTAGGGCAGATT
CTAATCTCTAATATTCCCTTCAACTCACTTTATTGGTGCCTTCTGGAGTGACATTAGA
AACTAAGACAAGAGGAAGATGAACAATGTTGAGGGATAGACAGCTGGATGCAATT
GGACCCAGATTCTAACACCGTCATGAAGCAAGTGATACACACAGTTATCAATAACGAGAAT
ATACACATAGATGGTTGAGTTATAAATGAACATTAAACGGCATTGTGGTTAGACAG
TGAGGAAGACGCCAGATAGACAAAGGGTAGGGGCTTGGTTAGATAATGAGAAGTTGAAG
AGGTGTAATAACTTAAATCTCTCTTGACTATTGATTGTGCTAAGAGTTTCTTATCTTA
CAGTCGGCCAGTTGGGTCAAAGATGGTGTGATTGGATGTGCTTGTGTTCTGCGATG
GCTGATTTAGAGTCAGTTACTTCAGATGAATGAAGTTCCCCGATTCTTATGTTAAGTT
TGTTTACCTACGCATGAAGACATCACCAGCAGGGTCGTCTTATTCTAGTAGCTTATT
TACAGCAAGCTTGTAACTGATGCTGAATTGCTGCGCTCTGTAGAACACACAGCATCTATGT
TTGCTTGTGCTTCTTAGTAGACTGCGGATGTGATGGTTGGTTACCTGGTATGCTGACGAA
GAATTGTTGACGTGGTGGTTGCCTTGATGGTTGCTGACTTGGTTGGATACTGA
TTAAGGTGACTCTGCTGGGAGGCTTGGATTCTGGGCCGGTCTTGTCTCCTGTCT
AGGGTGGCGATTATTCCCAACCCACTGTTCCATTACACTCAAAACCTGCTATCAATT
ACAG

DOMAIN 1A-3 (3rd part of domain a)

ATATTCAATGTCAAACCTGGAATACACCCCTACGACCCCATCTTCTCCACCACTC
CAACGTTGACGCCCTTCTGCCATCTGGCAGCGCTTCAGGAACCTGCGAGGAAAGAAC
CAATGCAATGGACTGTGACATGAACCTCGCTCACCAGCAACTCCAACCCCTCAACAGGG
CAGCAATCCAGTCCAGCTCACAAAGGACCACTCGACACCTGCTGACCTTTGATTACAA
ACAACCTGGATACAG

Fig. 4e

INTRON 1A-3/1A-4 (SEQ ID NO:113)

GTGAGACATTATTACACTTCTATTTAGTAGTGGGGCGGGATAGCTCAGGTGGTAGAGCG
 TCGGCCTTCAGCTCTAGTCTCGCCCACAAGAGCGCGCTGGCTAAAGGCCGGAGTTAGA
 TTCCCGCGGGCGGCAGGCAATATCTCCGAAGGGGAGAACAGTTCTCCAGTCGGTCAAATT
 GGGGTGCAATGTTGTAACCACTGAAATGCGTGCAGCACCAACCATCCAAATACCAGCCTG
 CCGCGCTGGTCTGACTACATAGTACCACTGAAATGCGTGCAGCACCAACCATCCAAATACCAGCCTG
 CAGCAGTAAATCTGACAGTCGCCATATAGCTGGATATTGCTGAGTGCACGTTAACGCC
 CAACTCACTCACTTATATTAGTATTCTATTAGTATCGACGCATGACCATGTGTGGTG
 GTCTACTCATCTCAACACGACCGATTAACGTTAAGAGCTGCCAACATGATTCTCTTCTC
 TCTTAGCCTCTTATGCCAAAAGCTATATATTAAATGTAGGACCCATATATTATTC
 CAG

DOMAIN 1A-4 (4th part of domain a)

CTACGACAGCTAAACCTGAATGGAATGACGCCAGAACAGCTGAAAACAGAACTAGACGA
 ACGCCACTCCAAAGAACGTGCGTTGCAAGCTTCCGACTCAGTGGCTTGCCCCCTCTGC
 CAACGTTGTTGTCTATGCATGTGTCCTGATGATGATCCACGCAGTGATGACTACTGCGA
 GAAAGCAGGCGACTTCTTCATTCTGGGGGTCAAAGCAGAACATGCCGTGGAGATTCTACAG
 ACCCTTCTTCTATGATGTAACCTGAAAGCGGTACATCACCTGGAGTCCCGCTAAGTGGCCA
 CTACTATGTGAAAACAGAACTCTCAGCGTGAATGGCACAGCACTTCACCTGATCTTCT
 TCCTCAACCAACTGTTGCCCTACCGACCTGGAAAGGTACCTTGACC

INTRON 1A-4/1B (SEQ ID NO:114)

GTAAGTTGATTGTCTTAATATTGTTAATTTCAGAAATTGATTAAATTGTGTA
 ATAACAGTACACATTTCAGCAACAGCAGTCATTATTGTGTGAAGATGTCAAACCAAG
 AAAGGTTCAATCGTAAAACAAAAACAATTCTCTATCTGTATAACCCCTCAATACCAGTA
 TGATCACAAATCTAGGAAATATTACAATACTGCTTCATAGAGTAACGTGTTGTGGCA
 GAGCTGGATACGAAGTTCTGATAGTTCACAGCTACATGATAGTAAATGAACCTGTACAC
 ATCAACGGTTGATCATGAAAATTGTATGTGTGAAAGTGCTACCTGTATTAGTGAACGT
 GCTACCTGTATAACTGAAAGTGCTACCTGTATGACTGAAAGTGCTACCTGTATGCTGAA
 GTGCTACCTGTATTAGTGAACGTGCTACCTGTATAACTGAAAGTGCTACCTGTATGACTG
 AAAGTGCTACCTGTATTAGTGAACAGTGTACCTGTATGAGTGAACGTGCTACCTGTATAA
 CTGAAAGTGCTACCTGTATGACTGAAAGTGCTACCTGTATTAGTGAACAGTGTGCTGCCTGTA
 TTAGTGAACAGTGTACCTGTATGACTGAGCGTGTACCTGTATGACTGAACGTGCTACCT
 GTATTAGTGAACAGTGTAACTGTATGAGTGAACAGTGTACCTGTATTAGTGAACAGTGTAA
 CTTGTATTAGTGAACAGTGTACATGTATGACTGAAAGTGCTACATGTATGAATGAGAGTG
 CTACCTGTGTGACTGAAAGTGCTACCTGTATTAGTGAACAGTGTACCTGTATGACTGAAC
 GTGCTACCTGTATTAGTGTACGTGACTGGTACCAACTGGATGTTCTCACTTCTTGGC
 GAATATCTGGGCTAAAACAGTTTCACTGATCATAGTCGTATCAGTTGATTGTATGT
 GCAGTGGAAATCATTTCGTCAAATAATCAAAACTGGTGTGAACGGCGTTCACGTTTA
 TGGTTGAAACAAATTCTGTAAGTAAAGATATTAGGGATATCTGTATGACATGAAC
 GAATTGCTTAAGGTTAGCATGCCATGACAAATTGCTGAATGTCTGAGGATTGGTGGAGCA
 ATAAATCATTATTAAGACAAAATCAGAAACGTCCTTCACTTTAACAGTGTATCTG
 TCTGAATGCCCTACTTTGGAAAGAGTATATGAATTATCGCAATATAAACGTTA
 AATGGCAAATGTCGGGCATATGTCAGGACATTATTACCGCAGTTATAGTCATATTAC
 GGGTCTAGGACAATTGTCACCCGACAATTGCCACCGACAATTGCCACCAAAATAA
 AATATACGTAACAGAAAACAATTGCTTCAGCCTTATTGAGTTAGATAATGACAT
 TTATGTTGATAAAATATGTCGTTGATAATAACAATAATATAATTACAATACT
 GCAATAGTACTATCAGTACTTATCATTTCACAGATTATAGATTCTAGAGTCG

Fig. 4f

ATGTTGTAGGCCAACACTCGTCGGTAGGCCGTTAGGTAGTTATCATTAGGGCTGAGTATT
 GCGCCAAATTCTGTATTGCTATATACTGCGATACACGGTTACCTGTTGCAATACGTA
 ACTTAGGCAAATATGACAGTTTCCATGATTATTTCACGTTCAATGCTAAAATGGT
 CTTATCTGTTATCTCCTTGAAGGTTAATAAAAACAATAAAACATAATCATTATTGAA
 AATTAATGAACAAAAGTAAAGCGCTCTCAGTTACCTAACCTAACCTTATTAATGAA
 GATTACTATCCAAGAATGTGAAATTCAAACACCTGGGATAACACTGCAAAACGACTG
 TTCATGGGACGGACATGAAAAAGGTGAGTCCCAGTAAACTGTTGAGAAAGTTCCSTAT
 ACTGTTGTCCCAGAAAAGGCTAAAGACCAGTGTACTAATCAATTATTCTATATTTCG
 ATTACTGTTCTCATATTGGGACAACGTGTGCAGATCGGTAGCATCCAAGCTCGTCAA
 CGGTTGATAAACCTTGTCAAATAACATGTTGTCACATCCAAGCTCACCTAACCTT
 GTCAATACCTGCATCTGAACAAATGTATATTAAAGACGATAGCATCCAAGCTCATCTTAA
 AAATGAATATTCTCTTTTACCAAAACATTATTGGTGACAGTTGTCCTCCSTAT
 TATAGAAAAAGAACTGGGTGCAATTGTCCTAGGTGGCAATTGTCGGATGGCAATTG
 CCGGGTGGCAATTGTCGGGTGGCAGTGTCCAGGGTGGCTATTGTCCTGTTCCATATT
 ACGTATCCCATTCTGCTCTGTAATTAAACTCACCTGCTTAAGGTAAGACGAC
 ATGTCACGTGAAACATCGTTGGGGCAAGGGCGGAATCCCTCGTTGAAAGTAAATGA
 ATACTGTACATAGAGATGCGTATCTGAACACTCTTATTAGCTTGATATTGTGCTTAATA
 TTACATGAATGTATTCAATATGTAATTATGTGTTCAAATGAATGGTGACTTGAATGGT
 TTTATTGCTTATATGCTACATCAACATGTGTTCTTCATTTCAG

DOMAIN 1B

CACCTGTGCATCATGCCACGATGACGATCTTATTGTTGAAAAAAATATAGATCATTGA
 CTCGTGAAGAGGAATACGAGCTAAGGATGGCTCTGGAGAGATTCCAGGCCACACATCCG
 TTGATGGGTACCAGGCTACAGTAGAGTACCATGGCCTCCTGCTCGTTGTCACGACCAG
 ATGCAAAAGTCAGGTTGCCTGTTGATGCATGGCATGGCATTCCCTCCCTCACTGGCACC
 GGCTGTTGTTACCCAGGTGGAAGATGCTCTTGTACGGCGTGGATCGCCTATCGGTGTC
 CTTATTGGGACTGGACAAAACCTATGACTCACCTTCCAGACTTGCATCAAATGAGACGT
 ACGTAGACCCGTATGGACATACACATCATAATCCATTCTCAATGCAAATATATCTTGT
 AGGAGGGACACCATCACGAGCAGGATGATAGATTGAAACTGTTGCCAGTCGCTT
 TTGGGGAGCATTCCCATCTGTTGATGAACTCTGTACGCATTGAGCAGGAAGATTCT
 GCGACTTGAGATTCAAGTTGAGTTAGTCCATAATTCTATTGCGTGGATAGGCGGTT
 CCGAAGATTACTCCATGCCACCCCTGCATTACACAGCCTTGACCCATTTCACCTTC
 ATCATTCCAATGTCGATCGTCTATGGCAATCTGGCAAGCTTCAAATCAGGAGACACA
 AGCCATATCAAGCCCACGTGACAGTCTGTGAAACAGTTGCCAATGAAGCCATTGCTT
 TCCCATCACCTTAAACAACAGAGAACACATAGTCATTGAGTCCGACTGACATT
 ATGACTACGAGGAAGTGCTGCACTACAGCTACGATGATCTAACGTTGGTGGATGAACC
 TTGAAGAAATAGAAGAACGCTATACATCTCAGACAAACAGCATGAACGAGTCTCGCGGGAT
 TTCTCCTGCTGGAATAGGAACATCTGCACTTGTGACATTTCATAAAATAACCGGGGA
 ACCAACCAACTCAAAGCTGGAGATATTGCCATTCTGGTGGTCCAAGGAATGCCTGGG
 CGTTGACCGCTTGTATAAGGTCGAAATAACTGACTCATTGAAGAACACTTCTCGATG
 TCGATGGAGATTATGAAGTCACTTAAAATTGATATGCACGGAAACGCTCTGATA
 CGGACCTGATTCCACACGCAGCTGTTCTGAGCCAGCTCACC

INTRON 1B/1C (SEQ ID NO:115)

GTAAGTAAATTACAAAATTGGTGTCTCTAACTATCCTAAGTATTCAATCGTTAGCGT
 GTACCTATCTGCATAATGCAATACCCCTGACTCCATATAAGTATAGTATATTACTCTGGT
 CGAAAACAAACAAATTGAAAACAAGAGTGGACGTGCTGTTATGATTCTTTCTTCAATT
 GGTCGTTGTGAATGCCACAGCCAGCAATTCCAGATATAGCGACGGTCTATGAATAC
 TCCAGTCTGGACCAGACAATCGTGTGGAATGGTTAGGCACATTATATCAAATTCTTGT

Fig. 4g

TGAAGATATGAGTTATGAGGTACAATGTTGTCAGTACGGCTGACAGCTC
ATTTCATGACTGAAATCTCTCAACGCCGTTAGCAATAATAGGCTCAGTAGTATTCAAC
CAATTACAATCAGTAGAAAATTCTCTATACTATTCTATGTTGCATCCTGATATCCCTAT
GCAAAAATTAGTCATCTAATATAATCATTTCGATAAATACTTGGCAAACAAATCAAT
GTAACATCTATTTCAG

DOMAIN 1C

CTACCTTGAGGATGAAAAGCACAGCTACGAATCAGAAAAATGTCGACAGCTGACTC
CTGAAGAAACAAATGAACCGCTGAGCTTCTGAAAATGATCATACTGCAG
GTGGATTCAATCAGCTGGCGCCTCCATGGAGAGCCTAAATGGTGCCTAATCCTGAAG
CGGAGCACAAGGTTGCATGCTGTGTCATGGCATGGCTGTTCCCTATTGGCACAGGC
TTCTGCTCTCCAGGGCGAGAATGCTCTAGAAAGCATGGGTACAGTGGTGCCTACCAT
ACTGGGATTGGACTCGCCCCCTTCCCAACTTCCTGATCTGGTTAGTCATGAGCAGTATA
CAGATCCTCCGACCACAGTGAAGCATAACCCGTGTTCAATGGCCACATCGATACAG
TAAATCAGGATACCACCAAGAGCGTACGGGAGGATCTTATCAACAAACCTGAATTGGAC
ATTCACGGATATTGCTCAACAAGTCCCTTAGCATTAGAACAGATGACTTCTGTTGCT
TTGAAGTGCAGTATGAGATTCCCATAATTATCCATGCACCTGTAGGAGGAACCGACG
CTTATGGCATGGCATCGCTGAGATATAAGCATAACGATCCAATCTTCTTCATCATT
CAAACACCGACAGGATCTGGCTATTGGCAATCCCTGCAAAAATACAGAGGCAAACCGT
ACAACACTGCAACTGCGCCATAGAACATCTATGAGAAGGCCCCTGCAACCATTGGACTAA
GCAGTGCCTTAACCCGTGACAGAACATCACCAGAGAGCATGCTATCCGTTGATGTCTTC
ACTATAGAGATAACCTCATTACGTATATGATAACCCGAAATTAAAGGTTGTCGATT
CACAACCTGATAGAGAGCTGGAAAAATCAAGAGTCACGAAAGAGTATTGCTGGATTCT
TGCTGTCGGGATTAAAAATCTGCTCTGAAATTGAAAGTTGACTCCACCTGATA
ATTGTCATAAAGCAGGGAGTTTATCTACTCGGGACGAAAACGAGATGGCTGGCCT
ATGACCGACTTTCAAGTATGATAATTACTCAGGTTCTGAAAGCAAACCATCTACACTTCT
ATGATCATCTCTCATTGCTACGAAGTCTTGTATCTAAAGGAGTGAGTTGGAACTG
ACCTGTTCCACACTGCAAATGTGGTACATGATTCCGGCACAG

INTRON 1C/1D (SEQ ID NO:116)

GTACGTGGATTGATTACATAGCAATGCTATATGATTCAGTAATTACAAACCTCAAGTCA
TGTAGCCGTTTAGATTGCATTACATCAAACAGCATTGGATTAAATTGGGGATTGTCCA
GGCCGCATTATGTCATTCCGAAAATAGTTGTGTCAGTGTCCACGTTAAAATTAAA
CCATTAAATCATATTAGGGATAATTAAATAGATGTTAGTGTCTTATTGATATTGT
TACAGTGGACAGTCACCAAGGACATATTACTCTATAGATAACACAAACACCAATTAAAA
CCCTGCTTGAAAGTCTAACTTTCCCCACAG

DOMAIN 1D

GCACCCGTGATCGTATAACTACGTTGAAGAAGTTACTGGGCCAGTCATATCAGGAAGA
ATTTGAACGACCTCAATACCGGAGAAATGAAAGCCTTAGAGCTGCTTCCTGCATATT
AGGACGACGGAACATATGAATCTATTGCCAGTACCATGGCAAACCCAGGCAAATGTCAAT
TGAATGATCATAATATTGCGTGTGTCATGGTATGCCTACCTCCCCAGTGGCACA
GACTGTATGTGGTTCAGGTGGAGAATGCTCTCCTAAACAGGGATCTGGTGTGGCTGTC
CTTACTGGGAGTGGACTGCTCCATAGACCATCTACCTCATTGATGATGCAACAT
ACTTCATCTCCGACAAACAGCGGTACGACCCCTAACCCCTTCTCAGGGAAAGGTTACTT
TTGAAAACGCAAGTCACAACAAGGGACCCACAAGCCGGCTTCAACTCAGATTATGT
ATGAGAATGTTTACTGCACTGGAGCAGGAAATTATTGTGACTTGAATTCACTGTTG
AGCTTGTTCATAACGCACTTCATTCCATGCTGGAGGTAAAGGGCAGTACTCCATGTCCT

Fig. 4h

CCCTGGACTATTCTCGTTGATCCCCTTCTTCACATCATGCCAACACGGACAGAC
 TGTGGCAACTGGCAGGAACACTAAAGATTCCGAGAACTGCCTATGAAGAAGCGAACT
 GTGCAATCACCTCATGCATCAACCACACTGAAGCGTTAGTGATCCACATGAGAACATCAGC
 ACAATGTCACTTGAAATACTCAAAACACAGGACGGATTCGACTACCAGAACCACTTCG
 GATACAAGTATGACAACCTTGAGTTCCATCACTTATCTATCCCAAGTCTTGATGCTACCC
 TGAAGCAAAGGAGAAATCACGACAGAGTGTGCGGGCTTCTTCTTCATAACATAGGAA
 CTTCTGCTGACATAACTATCACATATGTCTGCTGACGGACGGCGTGGCAATGACTGCA
 GTCATGAGGCGGAAACATTCTATATCCTCGGAGGCACAGAGATGCCTTTATCTTG
 ACCGTTGTATAAATTGAAATCACCAAACCACTGCAACAGTTAGGAGTCAAGCTGCATG
 GTGGAGTTTCGAACTGGAGCTTGAGATCAAGGCATACAACGGTCTATCTGGATCCCC
 ATACCTTGATCCAACATCATCTTGAAACCTGGAACAG

INTRON 1D/1E (SEQ ID NO:117)

GTAATGCCATCTTAATACAGTCGTTGTTAAATTATATGTTGTTACAACACCATA
 CCTTGAATTGAGGTAATACATCACTTGATATTGATAATGTAATGGTAATTGTTCTTGT
 GTAAAACCGTTCTGGGTGTTATTCACTATCCACCTGGTGGATAGTGAGTAAACACAT
 TCGGTTAATATGGGTATCTAATGGACAGTGAAGTGTGCTGGCTAGGCAGATACTGGT
 TTCTGTGAATGGAGGTAGTAGAAAGGGTTTGATGATTGCAG

DOMAIN 1E

ATACCCATATCTGGACCACGACCATGAGGAAGAGATACTTGTCAAGGAAGAAATATAATTG
 ATTTGAGCCCAAGGGAGAGGTTCTCTAGTCAGTCAGGTTCAAGCTTGCACAAAGAACATGATC
 GCTCCGCTGATGGGTACCAAGCCATTGCCCTTTCCATGCCCTGCCACACTCTGCCCA
 ATCCATCTGCAGCTCACCGTTATGCTGCTGTCCATGGCATGGCTACATTCCCCAGT
 GGCACAGACTGTACACTGTTCAAGGATGCCCTGAGGAGACATGGTCACTTGTG
 GTATTCTTACTGGGACTGGACAAAACCAAGTCACGAGTTACCCGAGCTTCTTCTTCAG
 CAACATTTATCCAAATCCGAATATTAATATTCAAATCCATTCTGGGCTGACA
 TAGAATTGAGGACCGGGCGTTCATACAGAGAGGCACATAACTGAGCCCTGTT
 ACAGTGGGATCATGACGGATACCACAACGGTTCTCGAAACTGTTCTTGTGTTGG
 AACAGGAAGATTACTGCGATTGAAATACAATTGAGATAGCCATAATGGCATCCACA
 CATGGATTGGGAGCGCAGTATATGGCATGGACACCTTCACTATGCATCATATGATC
 CAATTCTACATCCACCATTCACAGACGGACAGAACATGGCTATTGCAAGAGCTGC
 AGAAAGTACAGGGTCTATCTGGTCCGGAAAGCAAACGTGCCATTGAACATATGAGAACAC
 CCTTGAAGCCTTCAGCTTGGCCACCCATAATTGAATAGTCATACGCAAGAACATT
 CAAAGCCTGAGGACACGTTGACTATAAGAAGTTGGATACAGATATGATAGTCTGGAAT
 TGGAGGGCGATCAATTCTCGCATTGATGAACCTATCCAGCAGAGACAGGGAGAACACA
 GAACATTGAGGTTCTCTAAAGGTTGGTACATCCGATCTGTGTCATTGCAAG
 TTTGCAGAGTTGATCACACCTGAAAGATGCGGGCTATTCACTATTCTGGAGGATCAG
 CCGAAATGCCATGGCATTGACAGGCTTATAAGTATGACATTACTAAACTCTCAGC
 ACATGAACCTGAGGCACGGACACTTCTCTATAGACGTAACATCACGTCTAACATG
 GAACAGTACTCTGGAGACCTCATTCAAGACGCCCTCATTATATTGTACCTGGACGCC

INTRON 1E/1F-1 (SEQ ID NO:118)

GTGAGTACCTGTTGCACTAAGACTTCTGTAGGCTAAAGTGAAGAAATATCAATTAAT
 TTCAATTCAACCAAACCTGAAAACGGTACCTATATAGGTTAATTGCTACAGTAAA
 CTGAACATACACATTCATGAAATGATCTCAATATTTCCACCAACAG

Fig. 4i

DOMAIN 1F-1 (1st part of domain f)

ATAAACTCAACTCACGGAAACATAACACCTAACAGAGTCGCCATGAGCTAAGTAGCCTTA
 GTTCCCGTGACATAGCAAGCTTGAAGGCAGCTTGACAAGCCTCAACATGATAATGGGA
 CTGATGGTTATCAAGCTATTGCTGCCTCCATGGCGTCTGCAGTGCACAGGCCAT
 CTGGACGTGAG

INTRON 1F-1/1F-2 (SEQ ID NO:119)

GTAAATTTACAGAGCTTATGAAGTGTGTCAGAGTGAAGAGACCAAGATAACTTATAC
 CAAAAACTAGCTAGCAACAGACGATTCACTTGGACACTTGTATTATACGTTGG
 ATCCAAGGTAAACGGAAACGTAACCGAGAATCAGTCCGTAAGTGAGTGAGTTG
 GGGCTTAACGTCGCACTCAGCAATACCCAGCTATGTGGCGACTCTCAGATTACTGCTG
 GAGGAGAACCTACATAGCCGGTTAACCCGTGTTATGTAGTAAGACCAGCGCGGCAT
 GGCTGGTATCTGACGGACGAAGGGTGGCGCTGCACGTATTCCAGTGGTACAACACTGCAC
 CCCAATTTACCGACCAGGAGAACTGATCTCCCTCGGAGATATCGCCTGCCTCACGG
 GATTCGAACTCGGTACCTCAAGCCAGCGCGCTCTAGCGGGGGCGATTAGAGGTTNA
 GGCCGACGGCTCTACCACCTTAACATCCCCCGGCCACTCCTGACGGAAATGTTATA
 ATTCAGCCTTGTGTTCTATTAAACACTCTTGGCAGATTTCATAGATAATGGATTCA
 CATGTAGACAGTCTCCATTGTTGTACTGGTAGTCAAGAGTTAGAATCTGAATAACATT
 TCCAAGATGGATCAAGGAAAACAATAATTACTTGATGTTGCAG

DOMAIN 1F-2 (2nd part of domain f)

ATCGCCTGTTGCATCCACGGCATGGCAGCTTCCACTGGCACCGGTTGTACACTCTG
 CAGTTGGAGCAAGCGCTGCGCAGACACGGGTCCAGTGTGCTGTTCCATACTGGACTGG
 ACCAAGCCAATCACCAGAACTGCCACACATTCTGACAGACGGAGAATATTATGACGTTGG
 CAAAATGCCGCTTGGCCAATCCGTTGCAAGAGGTTATGTGAAAATTAAAGATGCATTT
 ACGGTGAGAAATGTCCAGGAAAGTCTGTTCAAATGTCAAGTTGGAAAGCACTCGCTT
 CTGTTGACCAGGCTTGTGCTCTGAACAAACTGACTACTGTGACTTCGAAGTTCAG
 TTTGAAGTGATGCATAACACGATCCATTATCTGTAGGAGGGCGTCAAACGTACGCCCTC
 TCCTCTCTCGAGTATTCTCATACGATCCAATCTCTTATTCAACCACCGTTGTTGAC
 AAAATATGGGCTGTATGGCAAGAACTGCAAAGCAGGAGACATCTACAGTTAGAACAGCT
 GATTGTGCTGGGCCTCATGGTCAGGCAATGAGGCCTTCAACAAGGATTCAACAC
 AACTCGTTACCAAGAACGACCGCAGTCCCTAACAGTATTGATTATGAAGATCTGGC
 TATAACTATGACAACCTTGAAATCAGTGGTTAAACTTAAATGAGATCGAGGCGTTAATA
 GCAAAACGCAAGTCACATGCTAGAGTCTTGTGGGTTCTGTTGGATTAGGAAC
 TCGGCTGATATACATCTGAAATTGCAAGACATCGAAAAGTCCATGATGCTGGTGTG
 ATTTCATCCTGGAGGTTCTGCAGAGATGCATTGGCATAACACGCCCTACAAGTAT
 GACATTACAGAACGATTGCAAGGAAATTGACATCAACCTGAAGATGTTCCATGCTGAT
 GAACCATTTCTGAGGCTGTCGGTTGTGAATGGAACGTGCAATTCCATCGTCT
 CATCTCACCAAGCCAACGATAATCTATGAACCAGGCGAAG

INTRON 1F-2/1G-1 (SEQ ID NO:120)

GTGAGATATGCAAATTGAATGTTGCCAGATGCGTTGTTACATTATGCTTGGAA
 TTGTCCTGAACGAATACAGTGGAAATAACCAAAAGCTGAAAAATAAAAGATATATACTTC
 ATTCTGAATTGTCAGTATTGCTGACCCAAAAACGTTATCCATGTCGACACTATATT
 GCCTTCTGAATCTGAGACTGCGTTATGTTCTAACATCACGAAATATGGTATACAGGT
 TGTGTATCTGAGAATACCCAAGGCAGAATTAAAGGGTACACCCCTGTTAACACAG

Fig. 4j

DOMAIN 1G-1 (1st part of domain g)

```

ATCACCATGACGACCATCAGTCGGAAAGCATAGCAGGATCCGGGTCCGCAAGGACGTGA
ACACCTTGACTAAGGCTGAGACCGACAACCTGAGGGAGGCCTGTTGCTCATGGCAG
ACCACGGTCCAATGGCTTCAAGCTATTGCTGCTTCCATGGAAAACCAGCTTGTGTC
CCATGCCTGATGCCACAACTACTCATGTTACTCACG

```

INTRON 1G-1/1G-2 (SEQ ID NO:121)

```

GTAAGTTGTGTTGGTTAGTGGTGCATGTTGCCATATCGATAGTATCAGTGTGG
TAACATCTGGTTCTAGTCATTCAGTCACCTTATCAGAACGCTGTTGCTCTCGTCTAC
AATAGTGACGTCTTCAGTTAGAACCGTGTACATCCGGTTATATTGGTCTCCAGCAA
CCCGTGCTTGTGAGGGCCACTGATGGAACGGGTGGTCAGACTCGCTCACTTAGTT
GACACATGTCAATTGCGAAGATCGATGCTGAGGGTGTAAACATTGGATTGTCTGGTCCA
GAUTCGATTATTACAGACAGCCGCGATGTACCTGGAATATTGCTGAGTGCAGCGTTAAA
CAACAAACTAGTCAGACTAATCTTCACTGTTATAATGATGGCTCGAACCTAGCACTCA
TGTCCAAGTTGGCAACATCTGGAAGGGAAATTCAAATGAAAAGAACATCTTCACGT
CTATTGGTATCAGCCTGGAGAACATGATGTTACGGCGTTACTTCCTTACCT
GTTTACTTGTCCCACGTTCTTCATATTAAAGAGTATTGGGTATTAGAGCTTGGT
GCTGTTACAATGCTACTCACTGTTAGTGCAGCGCCGACCGCGCTTGTACACATTAAGT
TTTGGTGTGGTTGGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
TGTGTGTGTGTGTGTATCTATGTCTATGTGTCTGTGTCTGTGTCTGTCTATGTGTGTG
TGTGTCTGTGTCTATGTGTGTCTGCGTGTGTCTGTGTCTGTGTCTGTGTCTGTGTCTA
TGTGTGTGTGTGTCTGTGTCTGTGTCTGTGTCTGTGTCTGTGTCTGTGTCTGTGTCTA
TGTGTCTATGTGTGTGACATGCAATACATGCTGTGATACTCACTAGCTGCGTCTATCGAC
CAG

```

DOMAIN 1G-2 (2nd part of domain g)

```

GCATGGCTACCTCCCACACTGGCATGCCCTACACCAAGCAGATGGAGGATGCAATGA
GGCGCATGGGTCTCATGTCGGCCTGCCACTGGGACTGGACTGCTGCCCTACCCACC
TGCCAAACACTGGTCACCGACACGGACAACAACCCCTCAACAT

```

INTRON 1G-2/1G-3 (SEQ ID NO:122)

```

GTAAGAGCGGGTAGGGATGGGTGGTAGGGGTGGTTCTATTACTCCGCTTCA
CTTGTATGAAATGGATAACCTGGCTGCATCCAAATTGCGTATCGATTCTCTTCGATT
CACTCGTGCATTAGACTGCCTTATTTACTATAGTAGTTAGAACATGTTGCTCAGTGCAGCG
TTAAACAACTAATACACAAAACCGCATTTGTTTATATGGTCACTACTGTTATCAGC
TATATGTATGTTCCGACTCACTGGTGGTGCCTACATTCTACTGTCACACTGAGAGCCA
ATGTTCTCAGATGTGAAATGTTGAAAGCCGTTCTACATAATATTGCAAGGAATACCA
TTGTAGAATGTAGTCAAACAGGTAACAATCTGTTAGTGAGGCCAGTCGAGGTTGCGTTG
TAGGGTAGTCCACAGGTAGGCAGTCCATAAGCATAGTTTAAGCATTTAGATCAT
CTATAATTAAACCACATGGTTAGCCGCTATGTTAGTTAATCCAGTATAAGTTAGAACTG
TTATATTCGAAGGGAGTGAGTAAATCCTTATTCCTGACTACCATTAAAGATTCC
CAATGACTCCATTCAACTCCTAACATCACTGCTCTTCAACAG

```

DOMAIN 1G-3 (3rd part of domain g)

```

GGACACATTGATTATCTCAATGTCAGCACAACTCGATCTCCCCGAGACATGCTGTTCAAC
GACCCCGAGCATGGATCAGAGTCGTTCTACAGACAAGTCCTTAGCTCTGGAAACAA

```

Fig. 4k

ACTGATTCTGAAATCGAAGTTAGTTGAGATAACCCACAATGCCATCCATTCTGG
 ACAGGTGGCACAGCCCTACGGAATGTCCACTCTGACTTCAGTGCCTACGATCCTCTC
 TTCTGGCTTCAACCACCAACACCGACAGAACATCTGGCTGTCTGGCAAGCTTGCAAGAA
 TACAGAGGACTTCCATACAACCATGCCAATTGTGAGATCCAGGCAATGAAAACGCCCTG
 AGGCCTTCAGTGAATCAACCACAAACCCAGTCACAAAGGCTAACGCGAAGGCCATTA
 GATGTGTTCGAGTATAATCGTTGAGCTTCAGTACGACAACCTCATCTTCATGGATAC
 AGTATTCCGGAACCTGATCGCGTCTGAAGAAAGAAAGGAGGAGACAGAATATTGCT
 GCCTTCTTCAGTGAATCAAGCGTAGTGCTGATGTAGTGTGACATATGCCAGCCA
 GAACACGAATGTGTGTCGAGGGACTTTGCGATTTGGGAGGGAGCTAGAAATGCC
 TGGTCCTTCGACAGACTGTTCCGCTATGATATACCAAGGTGATGAAGCAGCTACACCTG
 AGGCATGACTCTGACTTTACCTTCAGGGTGAAGATTGTCGGCACCGACGACCACGAGCTT
 CCTTCAGACAGTGTCAAAGCACCAACTATTGAATTGAACCGGGCG

INTRON 1G-3/1H (SEQ ID NO:123)

GTGAGTACGACAGGCATTCTAGTAAAAACCTACTTTGGTAAAAGGTTCGAGAAATCAC
 TTGAAGCAACACATGATTTGTAACGCCATTACACGTGAACATGTCACACCCGGTGT
 GCCGTTAATGGACATGCCCTGTTAATGAAAGGGTAAGTACATGTGTATGGGATGGG
 ATGGGAGCCACCTGTCCAATTCTAGGTCCCTAGGATCCCAGTTGCGTAGGAATCCCC
 TGATTAATGCCCTGTGAATTCCCTGGAATTGTCCTGGCCAAATTTCACAAACCCGC
 CCCGATATACTTGGAAATAATTGGCCTAAGGGTGGGCTTTAAGGACCAAGAACCCA
 ACCTAAACCCCAACCCATTTCACCCATTCCAGGTTGTTACCAAATAAAAG
 GTTCCACTTGAGGAAACCTTAAGGGTTCTTCAAGGGCTTTCTTTCTGGGA
 ATTCCAATTCCGGGGAAACAAATACATATATTACAGACCTTGTCAAATTATATA
 ATTCGACTCATGTCATAGGTTGTTCTACACAG

DOMAIN 1H

TGCACAGAGGCGAAACACAGAAGATGAACACCATGATGACAGACTCGCAGATGTCTGA
 TCAGGAAAGAAGTTGACTTCCTCTCCCTGCAAGAGGCCACGCAATTAGGATGCACTGT
 ACAAGCTCCAGAATGACGACAGTAAAGGGGCTTGAGGCCATAGCTGGCTATCAGGGT
 ATCCTAATATGTGTCCAGAAAGAGGTACCGACAAGTATCCCTGCTGTCCACGGAATGC
 CCGTGTCCCCACTGGCACCGCCTGCATACCATTAGTGGAGAGAGCTCTGAAAAACC
 ATGGCTCTCCAATGGCATTCTACTGGGATTGGACAAAGAAGATGTCGAGTCTTCAT
 CTTCTTGAGATTCCAGCAACAACACCCTTCTACAAATATTACATCCGGGGCGTGC
 AGCACGAAACAACCAGGGACATTATCAGAGACTTTAATCAAACCAAGTTGGTGAAT
 TTGATTACCTATATTACCTAATCTGCAAGTCCTGGAGGAAACTCGTACTGTGACTTTG
 AAGTCAGTATGAGATCCTCCATAACGCCGTCCACTCCTGGCTTGAGGAAGTGGAAAGT
 ATTCCATGTCTACCCTGGAGCATTGGCCTTGACCCCTGCTTCATGATTACCAACTCGA
 GTTTGGATAGAATCTGGATCCTTGGCAGAAGTGCAGAAAGATAAGAATGAAGCCTACT
 ACGCATTGGATTGTGCTGGCAGACAGACTATGAAAGACCCCTGCATCCCTCAACTACG
 AAACCGTTAATGAAGATGAATTCAACCCGATCAACTCTTCCAAGCATACTGTTGACC
 ACTACAGGTTCAACTATGAATAACGATAACATGAGAATCAGGGTCAGGACATACATGAAC
 TTGAAGAGGTAATTCAAGGAAATTAGAAACAAAGATCGCATATTGCTGGTTTGT
 CGGGCTTACGGATATCAGCTACAGTGAAGTATTCAATTGAAACAGATAACAAGTC
 ACGAAGAATATGCAGGAGAATTGCAAGTTGGAGGAGGTGAGAAGGAGATGCCGTGGCAT
 ATGAAAGAATGCTGAAATTGGACATCTCGATGCTGTACACAAGCTTCACGTGAAAGATG
 AAGACATCCGTTTAGAGTGGTTACTGCCTACAACGGTGACGTTGTTACCAACCAAGGC
 TGTCTAGCCATTGTCACCGTCCACCGTCCAGCCATGTGGCTACGACATCTGGTAATCC
 CAGTAGGTGCGGGCATGACCTCGCCTAAAGTCGTAGTAAAGAGCGGCACCAAAGTCG

Fig. 4I

AGTTTACACCAATAGATTGTCGGTGAACAAAGCAATGGTGGAGCTGGGCAGCTATACTG
 CTATGGCTAAATGCATCGTCCCCCTTCTCTTACCACGGCTTGAACTGGACAAAGTCT
 ACAGCGTCGATCACGGAGACTACTACATTGCTGCAGGTACCCACCGCGTTGTGAGCAGA
ACCTCAGGCTCCACATCCACGTGGAACACGAGTAG

3' UTR

TTCACAG

INTRON 3'UTR (SEQ ID NO:124)

GTGAGGAGAAGGCCAGGCTAGCAGGGCAATGGATGAAGGAAATAGGGCAAAGGGAAT
 AGCAGTTACACCATCGACATTCCAACCTCCTCAGAAACTAATATATAGCCTTAATACAA
 CCAGCCAAGACTCAACGGGCAGCCGGGGATTGGTGGTCGCTGTTCAGACCA
 GGGTCAAATATCAGTGCACAAATCAACATGTTGCGTGTACAGACACTGACACAGCAGTC
 ATTGAACCTGCAGACCCATAACAGGAAAATGGGGCAGATACGATCAAAGACAGTGTAAAA
 TAGGGATAAGTAGGCATATGCAACCACCTGATGAAATGAAAAGGGTAAGTTAACCC
 CGGCTACAAAGGTCCAATGGTCTTAACCCAGCTACGCTATCCCTCTAATTCAGTA
 TTGAGCTGATTCTGTCAGTTCATGTAAACTGTATACTTCTGTATTATTACAG

3' UTR

GTTGCTATGCCGACTGCGCTATATTGGTGAACGAGACGATGAGGACATCTCTGAAAGAGT
 TCGCCAAGTGATGTGTAGGTACGGAAGTATTGTTGAGCTAACATATGATGATTCAAA
 ATGACTTGGCGCTCTAGGACAAAGACATAATTCATCAGCACCCCTGTGCACCAACTCTTG
 TTTGCTGCAAACGTCTGACAAGCGACACGTCAATCAACAAGCTGTTCAAACACTCAAGTGG
 TGTAACTAGAATCGTGGGCCATCGTTACAAAGTATTGACAGATGTCACACATGATGGC
 GAGAAACACTTTAGAACTTTAATGACCTAGAGTGACTTGTAAATATGAAATATATTCT
 TCAAAGACTCAGCTGAACATTGTTGGATAACACATCAATTCCCTCAACAAATGCTTTA
 TCTTCACATGGATGTATGTAATGTGGCCGGCAATAAGTATATATGTAT

Fig. 5aPrimary structure of the HtH1 proteinSIGNAL PEPTIDE

LVQFLLVALVVGAGA

DOMAIN A

DNVVRKDVSHTVDEVQALHGALHDVTASTGPLSFEDITSYHAAPASCDYKGRKIACCVHGMPSFP
 FWHRAYVVQAERALLSKRKTVGMPYWDWTQLTLHPLSLVTEPIYIDSKGKAQTNWYRGEIAFIN
 KKTARAVDDRLFEKVEPGHYTHLMETVLDALEQDEFCKFEIQFELAHNAIHVLVGGKFEYSMSNLE
 YTSYDPIFFLHHSNVDRLFQIWRQLQELRGKNPNAMECAHELHQQLQPFNRDSNPVQLTKDHSTP
 ADLFDYKQLGYSYDSLNLNGMTPEQLKTELDERHSKERAFAFRLSGFGGSANVVYACVPDDDP
 SDDYCEKAGDFFILGGQSEMPWRFYRPFFYDVTEAVHHLGVPLSGHYYVKTELSVNGTALSPL
 PQPTVAYRPGK

DOMAIN B

GHLDPVHHRHDDDLIVRKNIIDHLTREEEYELRMALERFQADTSVDGYQATVEYHGLPARCPRPDA
 KVRFACCMHGMASFPHWHLFVTQVEDALVRRGSPIGVPYWDWTPKPMTHLPDLASNETYVDPYGHT
 HHNPFFNANISFEEGHHTSRMIDSKLAFAPVAFGEHSHLFDGILYAFEQEDFCDFEIQFELVHNSI
 HAWIGGSEDYSMATLHYTAFDPIFYLHHSNVDRLWAIWQALQIRRHKPYQAHCAQSVEQLPMKPF
 FPSPLNNNEKTHSHSVPTDIYDYEELHYSYDDLTFGGMNLEEIEEAIHLRQQHERVFAGFLLAGI
 GTSALVDIFINKPGNQPLKAGDIAILGGAKEEMPWFDRLYKVEITDSLKTLSDLVDGDYEVTFKI
 DMHGNALTDLIPHAAVVSEPAH

DOMAIN C

PTFEDEKHSLRIRKNVDSLTPPEETNELRKALELLENDHTAGGFNQLGAFHGEPKWCPNPEAEHKVA
 CCVHGMAVFPWHRLLAQAEHALRKHGYSALPYWDWTRPLSQLPDLVSHEQYTDPSDHVKHNP
 WFNIGHIDTVNQDTTRSVREDLYQQPEFGHFTDIAQQVLLALEQDDFCSFEVQYEISHNFIHALVGG
 TDAYGMASLRYTAYDPIFFLHHSNTDRIWAIWQSLQKYRGKPYNTANCAIESMRRPLQFGLSSAI
 NPDRITREHAIPDFVFNYRDNLHYVYDTLEFNGLSISQLDRELEKIKSHERVFAGFLLSGIKKSAL
 VKFEVCTPPDNCHKAGEFYLLDENEMAWAYDRLFKYDITQVLEANHLHFYDHLFIRYEVFDLKGV
 SLGTDLFHTANVVHDSGT

DOMAIN D

GTRDRDNYVEEVGTGASHIRKNLNDLNTGEMESLRAAFLHIQDDGTYESIAQYHGKPGKCQLNDHNI
 ACCVHGMPTFPQWHRLYVVQVENALLNRGSGVAVPYWEWTAPIDHLPFIDDATYFNSRQQRYDPN
 PFFRGKVTFENAVTTRDPQAGLFNSDYMENVLLALEQENYCDFEIQFELVHNALHSMLGGKGQYS
 MSSLDYSAFDPVFLHHANTDRLWAIWQELQRFREL PYEEANCAINLMHQPLKFSDPHENHDNVT
 LKYSKPQDGFDYQNHFGKYDNLEFHLSIPSLDATLQRRNHDRVAGFLLHNIGTSADITIYIC
 LPDGRGNDCSHEAGTFYILGGETEMPFIGDRLYKFEITKPLQQLGVKLHGGVFELELEIKAYNGS
 YLDPHTFDPTIIFEPGT

DOMAIN E

DTHILDHDHEEEILVRKNIIDLSPRERVSLVKALQRMKNDRSADGYQAIASFHALPPLCPNPSAAH
 RYACCVHGMATFPQWHRLYTVQVQDALRRHGSVGIYWDWTKPVNELPELLSATFYHPIRNINI
 SNPFLGADIEFEGPGVHTERHINTERLFHSGDHDGYHNWFFETVLFALEQEDYCDFEIQFEIAHNG

Fig. 5b

IHTWIGGSAYGMGHLHYASYDPIFYIHHSQTDRIWAIWQELQKYRGLSGSEANCAIEHMRTPLKP
 FSFGPPYNLNSHTQEYSKPEDTFDYKKFGYRYDSLELEGRSISRIDEIQQRQEKDRTFAGFLLKG
 FGTSAVSLQVCRVDHTCKDAGYFTILGGSAAEMPWFDRLYKYDITKTLHDMNLRHEDTFSIDVTI
 TSYNGTVLSGDLIQTPSIIFVPGR

DOMAIN F

HKLNSRKHTPNRVRHELSSLSSRDIASLKAALTSLOHDNGTDGYQAIAAFHGVPACHEPSGREIA
 CCIHGMAFPFWHRLYTLQLEQALRRHGSSAVPYWDWTkpITELPHILTGEYYDVWQNAVLANP
 FARGYVKIKDAFTVRNVQESLFKMSSFGKHSLLFDQALLALEQTDYCDFEVQFEVMHNTIHVLVGG
 RQTYAFSSLEYSSYDPPIFFIHHSFVDKIWAVWQELQSRRHLQFRADCAVGLMGQAMRPFNKDFNH
 NSFTKKHAVPNTVFDYEDLGNYDNLEISGLNLNEIEALIAKRKSHARVFAGFLLFGLGTSADIHL
 EICKTSENCHDAGVIFILGGSAAEMHWAYNRPLYKYDITEALQEFDINPEDVFHADEPFFLRLSVVAV
 NGTVIPSSHQPTIIYEPGE

DOMAIN G

DHHDDHQSGSIAGSGVRKVNTLTKAETDNLREALWGVMDHGPNQFQAIAAFHGKPALCPMPDGH
 NYSCCTHGMATFPFWHRLYTKQMEDAMRAHGSHVGLPYWDWTAAFTHLPTLVTDNNPFQHGHID
 YLNVSTTRSPRDMLFNDPEHGSESFFYRQVLLALEQTDFCFKFEVQFEITHNAIHSWTGGHSPYGMS
 TLDFTAYDPLFWLHHSNTDRIWAVWQALQEYRGLPYNHANCEIQAMKPLRPFSDDINHNPVTKAN
 AKPLDVFEYNRLSFQYDNLIFHGYSIPELDRVLEERKEEDRIFAAFLLSGIKRSADVVDICQPEH
 ECVFAGTFAILGGELEMPWSFDLFRYDITKVMQLHLRHDSDFTFRVKIVGTDDHELPDSVKAP
 TIEFEPG

DOMAIN H

VHRGGNHEDEHHDDRLLADVLIRKEVDFLSLQEANAIKDALKLQNDDSKGGFEAIAGYHGPNMCP
 ERGTDKYPCCVHGMPPVFPWHRLHTIQMERALKNHGSPMGIPYWDWTKKMSSLPSFFGDSSNNNP
 YKYYIRGVQHETTRDINQRLFNQTKFGEFDLYYLTLQVLEENSYCDFEVQYEILHNAVHSWLGTT
 GKYSMSTLEHSAFDPVFMIIHSSLDRIWILWQKLQKIRMKPYYALDCAGDRLMKDPLHPFNYETVN
 EDEFTRINSFPSILFDHYRFNYEYDNMRIRGQDIHELEEVIQELRNKDRIFAGFVLSGLRISATVK
 VFIHSKNNTSHEEYAGEFAVLGGEKEMPWAYERMLKLDISDAVHKLVKDEDIRFRVVVTAYNGDV
 VTTRLSQPFIVHRPAHVAHDILVIPVGAGHDLPPKVVVKSGTKVEFTPIDSSVNKAMVELGSYTAM
 AKCIVPPFSYHGFELDKVYSVDHGDDYIAAGTHALCEQNLRLHIHVEHE

Fig. 6aGenomic sequence of the HtH2 gene

DOMAIN 2A-1 (1st part of domain a)
[domain a, parts 1-4: SEQ ID NO:156]

GGTCTTCGTACTGGACTGGACGCAGCATCTGACTCAACTCCCAGATCTGGTGTAGACCCCTTG
TTTGTGACCCGGAAGGAGGAAAG

INTRON 2A-1/2A-2 (SEQ ID NO:125)

GTAAGGGATCTCAGATCCGTAGAGTGAGTGAGTGAGTGAGTCCCCAGCAACTGAAGCTAGG
CCGCCCTACTGGGATCACAGGAATGTATGTCAATGGTGAAGAAAGGAGCAGTGGTTACAACG
CCCGCTCAAAGTCATGGCAGTTCATAGCGCATTGTGCGCGTGTGTATCTGTGTGCGCGTGT
TGTGCTTGCCTGCGTGTGAGTGAGTCCGCTTGTGACTAGCACAGACTAACAGTGGTTCT
AGAGAGCCTACTGATAAAATGTTACATTAAGATCTTACAGTATACTGAGATTGAGCCAGACCA
GCGGAACACCAGGCAGGGTAACAACAAATAACGCCCTTCCACACAACCGACGCAGCCTAAAGTGGC
TCTGATAGGCTGATACCGGTGATTCTTAGAACTTGTAAATTGTGCTTGCATAATACATGTACT
TCAGTTAACTGTAATACAGCATAAGACTGGACCGGTGTTACGACGCAATGAGCAATAATTACT
ACGAAAAGATTGTTAGACATATTCAATAATTGTAACATTCAATTAAACATGAACACCACTGCAC
TCTCGTTGTGTCACGTATTCTATGCACTTCATGCATCTGTTAGCTCAGATATTTGATGTT
CAAGAGATTGTCAGAACGTATGGCTGGTGCCTGAAATTACATACAATGAATTCAAGGTGAAA
TACCTGGCGAGACAATAAGATCTTACTAGTGCTGCCACTTCAGTATGGTGTCCCCGATGGTGTCTG
GTGTATGGGTGTGTTGGCGTCAGTTGTTACTGGAAAAGTCAGCTCTAATTATGTCTTATGTGGT
TAAAGACCCCATAACCTAGATGTCTGGTTAACATGATAGTAACAGTCGGCTGTATAGCC
TGACGCTTAAACGTTAGATGAATAAGGACTATATTGTTGATACATTCTATAACCTCCTTC
TATATCATTAG

DOMAIN 2A-2 (2nd part of domain a)

GCCCCATGACAACGCATGGTATGTGGAAACATCAAGTTGAGAATAAGAAGACTGCAAGAGCTGTT
GACGATCGCCTTTGAGAAGGTTGGACCAGGAGAGAATACCGACTCTTGAAGGAATTCTCGAT
GCTCTTGAACAGGATGAATTCTGCAACTTCAGATCCAGTTGAGTTGGCTCACACGCTATCCAC
TACCTGGTTGGCGGCCGTACAC

INTRON 2A-2/2A-3 (SEQ ID NO:126)

GTGAGTCACGTTCTCTGATGGTCACGAGTCACGTTCTCTGATGGTCACGAGTCACGTTCTCTGATG
GTCACGAGTCACGTTCTCTGATGGTCACGAGTCACATTCTCTGATGGTCACGAGTCACATTCTCTG
TTGAGTGAAGTCTCAGTACCAATTATTCTCTTACCTTCTTAACCAGGGTTTCAGCGTGGATC
GTCTGAGAAGTTAGCGCAAATCTATATTGAAGTCATTTCATCATATAACCACGTTATATCCA
CGTGCAGAAAGTGTTCATTAATTATTTCATTTATGAAGGCTAAAAGAAAATATGTATTG
TTGGAAACTATATTGAGGTGAAGGCAACACGAGTGATTAATATTCTCAATATCAATGTACGCT
CTGTCAGCACCTGTTACCCAGGAACACACCTTACGCTACCAAAATATCAGCTGATGATTGCA
AGCGGACTATACCCCTACCAACTTGTGTTGTGTGATTTATGTGTGATGTGTGCGTGC
GTGTGTGTGTGTCCTACGTATGTTGATATTGTTCTGACTGTATATGTTGCTTACCAATTG
AAG

Fig. 6b

DOMAIN 2A-3 (3rd part of domain a)

GTACTCCATGTCTCATCTCGAGTACACCTCCTACGACCCCCCTTCTTCCTCCATCACTCCAACAC
CGACCGCATCTCGCCATCTGGCAACGTCTCAGGTACTCAGAGGAAGGACCCAACACCGCCGA
CTGCGCACACAACCTCATCCATGAGCCCATGGAACCCTCGTGGACTCGAACCCCTTGACCT
CACCAGGGAAAACCTCAAACCAATTGACAGCTTGATTATGCCACCTGGCTACCA

INTRON 2A-3/2A-4 (SEQ ID NO:127)

GTATGTATGATTCTAATAATGAATGTTTACCTCCGGTTAAACAATATTTAGTATTACGAAAG
GAGAAGTACCTCGAGAGGTCTAGGTCTCAGATGTTAGAAACCCATGAAGACAGGTATGCTTCTGA
AAAACAAAGTAACATCATGAGGCTAAAGTTAGTCAAAACCATCGTAGTTGAATCCAGCATGCA
AAGGGCCCTAACCCCTGTAGATGGCGCTGCTGAAACAGAGTAGTCTGTTCAAGGGTCAGTACTGTCC
CCACAAACATCATAGTCAGGGTCAGTACTGCCCCAAACATCATAGTCAGGGTCAGTACTGTCC
CCACAAACATCACAGTCAGGGTTAATTTGGATTCTGGTTCTGAATGCCAAGAACAGTCACGCC
TGACACTGGACCGAGGTTGCCGAGAAAGCTCGTATATTGCTGGAATACTGCCAGTAAACCATC
ATTTATTTAGGCTATTTATTACGAAAAATAATAATATGTATAGAAATGCATATGATCGCTGTTG
AATGTAAAATTAGAATGGGTTGGGAGTGTCACTATTTTCTCAAAATTCTATGTATTTAA
CCGATCGACGCTGAAGACAAACTACCGTTAACGGCAGTTCTATCTGATAGGGAAATTG
GTTGTTAACCAACGCTACATTGTGTCCAG

DOMAIN 2A-4 (4th part of domain a)

GTATGATGACTTGACCCCTGAACGGTATGACCCCAGAGGAATTGAACTCATATCTGCATGAACGGTC
AGGCAAGGAGGGGTGTTCGCAAGCTTCCGACTCTCAGGTTGGCGGCTCTGCTAACGTTGTTGT
CTACGCATGCCGTCCTGCCACGATGAAATGGCTGTCGATCAGTGCACAAAGCCGGCAGTTCTT
TGTGTTGGCGGACCCACCGAGATGCCCTGGAGGTTTACAGAGCATTCCACTTCGACGTCACCGA
CAGCATCGACAACATCGACAAGGACGCCACGCCACTATTATGTAAGGCGGAATTATTCAGTGT
AAATGGAAGTGCCTACCGAATGATCTCCTGCCTAACCCACCATCTCACACAGGCCAGCCGG
ACACGTTGATG

INTRON 2A-4/2B (SEQ ID NO:128)

GTAAATGCCATTGTATACATGCATTCTGGACTTGAGTGAGTGAGTGAGTGCGTATTCA
AGTGAGAGTGTGAGTGGGTATTAGGTCTGTGAGTGGGTGGTGAGTGGATGGGTGAGTAAGAGTGG
GTTGGTGAGAAAGTGAGTGAGTCACTGGTGGGTGCGTTAGTGGAAAGCGTGATTGAGTGAGTGG
GGTAGGTGAGTGAGTAATTGGTGGGGGGTGAGTGGTTAACGCTGTTCTGTTCAATCACA
CCACATGTTGCCAGCTTACTGTGCAGGACGAATCCAGGGTTGTTAAATTATGTTATATA
TAACGATGGACGTGCTGGATGTGGGAATGTGCAAGAGAATTATGCGGCTTGTGCTGCTCCGC
GTATTTATTGCACCGCGTGGTACCGGTTGATAAAAGTAGTTCAAAACATTCCAGCCATCTT
GTCTGTTGTGAAACCTACTCCAGGACCATCCATTCAATATGTGCTGCGTTATGGAGTTATAC
ATGTTAAACTGTAGAGCGCAGATGAGCACACTTGAGCATTCTCAGTAAATCAGAATGTGTATAT
TTCAAAATTACCAAATGCAATATCATCAAGCAAATTATGCAAGCTCTATAGTAACATCGGAGTCAA
TGGTCCAGTGTGCCCTCGGCTGCCATTCCGACCTCCCTGGCCAGAATACACCCGGTCAGGATCAG
TTATCCGTAGAAGGCACGGTGCAGGAATGAAAACATAAACACATAGTCGCTTAGTAGTATGCTGAT
TTAGGCACGCAAATCCGAATGTGAATTACTGTGAATTGCAATTACCTGTTACAG

Fig. 6c

DOMAIN 2B

AGGCCCCAGCTCCCTCCTCGGATGCTCACCTGCCGTAGGAAGGATATCAACCACGACACGCG
 AGGAGGTGTACGAGCTGCCAGAGCTATGGAGAGATTCCAGGCCACACATCCGTTGATGGTAC
 AGGCTACGGTTGAGTATCACGGCTTACCTGCTCGATGTCCATTCCCCGAGGCCACAAATAGGTTG
 CCTGTTGCATCCACGGCATGGCAGATTCCCTCATTGGCACAGACTGTTGTTACCCAGGTGGAAG
 ATGCACTGATCAGGCAGGGATCCCCATAGGGTCCCTACTGGGACTGGACTCAGCCTATGGCAC
 ATCTCCCAGGACTTGCAGACAACGCCACCTATAGAGATCCCACAGCGGAGACAGCAGACACAACC
 CGTTCCACGATGTTGAAGTTGCCTTGAAAATGGCGTACAGAACGTCACCCAGATAGTAGATTGT
 TTGAACAACCTCTATTGGCAAACATACGCGTCTCTCGACAGTATAGTCTATGCTTTGAGCAGG
 AGGACTCTGCGATTGAAAGTTCAATTGAGATGACCCATAATAATATTACGCCCTGGATTGGTG
 GCGCGGGAAAGTATTCCATGTCTCTACACTACACAGCCTCGACCCATATCCTCACCTCATC
 ACTCCAACACTGACCGTCTGGCAATTGGCAAGCGTTGCAGATACGAAGAAACAAACCGTATA
 AGGCTATTGTGCTTGGTCTGAGGAACGCCACTCAAACCTTCGCCTCAGTTCCACTGA
 ACAACAACGAAAAACCTACGAAAACCTCGGTGCCAACGTTACGACTACGAAGGAGTCCTTG
 GCTATACTTATGATGACCTCAACTCGGGGCATGGACCTGGTCAGCTTGAGGAATACATCCAGA
 GGCAGAGACAGAGAGACAGGACCTTGGCTGGCTTCTGTACATATTGGTACATCAGCGAATG
 TTGAAATCATTATAGACCATGGACTCTTCATACCTCCGTGGCACGTTGCTGTTCTGGCGGAG
 AGAAGGAGATGAAATGGGGATTGACCGTTGTACAAATATGAGATTACAGATGAACGTGAGGCAAC
 TTAATCTCCGTGCTGATGATGGTTTCAGCATCTCTGTTAAAGTAACGTGTTGATGGCAGTGAGC
 TGTCCCTGAACTCATCCACATGCTGCTATCATTGCAACGAAGCCATA

INTRON 2B/2C (SEQ ID NO:129)

GTAAGTAGCTACCTGTTATTCAATTTCGCTTGCCAATCAATTCACTCAGCTGAAATTCAA
 TAATTGTGTTTGCAATGGCTGAAAACCAATTGAACTCTTCTTTCTCAGGTGAACTCAAATA
 AATAATCACTAATTGTTATGCACGCCGTAGGGCATACATAACTATATCCACATCGGTATCTCAA
 ATGCAAACAAATTGTCTTATTCCGTTGGACAAGCAAACCCCTTCCTGTAATCTGCCTTGG
 CATCCACTGGAATTATGTTGACTGGTAATTGATACTGGCTCTCTGATAGAGTTAATATCT
 ATAGTTGAAATCTTATGATTTGCTATTATTCGACAGCATGCTATAGACACCCTAGACT
 ATTGTATAGCCACTGTATTGTTCCATTATTATAACAGAACATGGCTGTAATTGTTA
 TTTACCTCCAG

DOMAIN 2C

TTGACCATCAGGACCCCTCATCAGGACACAATCATCAGGAAAATGTTGATAATCTTACACCCGAGG
 AAATTAAATTCTCTGAGGAGGGCAATGGCAGACCTCAATCAGACAAACCGCCGGTGGATTCCAGC
 AAATTGCTGCTTTACGGGGAAACCAAATGGTGCCCAAGTCCGATGCTGAGAAGAAGTTCTCCT
 GCTGTGTCATGGAATGGCTGTCTCCCTCACTGGCACAGACTCCTGACCGTGCAAGGGAGAATG
 CCCTGAGAAAGCATGGATGTCTCGGAGCTCTCCCTACTGGGACTGGACTCGGCCCCCTGCTCACC
 TACCTGATTGGTAAGTCAGCAGAACTACACCGATGCCATATCCACCGTGGAAAGCCGAAACCCCT
 GGTACAGCGGCCATTGATACTGGTGTGACACAACAAGAGCTCCGTCAGAACACTGTATG
 AAGCTCCGGATTGGTCATTATACTGGGTCGTAAGCAAGTGCTCTGGCTTGGAGCAGGATG
 ACTTCTGTGATTTGAAGTCCAGTTGAGATAGCTCACAAATTGATCCACGCTCTGTCGGCGGAA
 GCGAGCCATATGGTATGGCGTACTCCGTTACACTACTTATGATCCAATTTCACCTCCATCATT
 CTAACACTGACAGACTCTGGCTATATGGCAGGCTCACAAAGTACAGGGCAAACCTTACAATT
 CCGCCAACGTGCCATTGCTTCTATGAGAAAACCCCTACAGCCCTTGGTCTGACTGATGAGATCA
 ACCCGGATGATGAGACAGACAGCATGCTGTTCCAGTGTCTTGATTACAAGAACAACTTCA
 ATTATGAATATGACACCCTGACTTCAACGGACTATCAATCTCCAGCTGGACCGTGAACGTGCA
 GGAGAAAGTCTCATGACAGAGTATTGCCGGATTGGTCTGACAGTGGTATTCAAGCAGTCTGCACTAG

Fig. 6d

TTAAATTCTTGTCTGCAAATCAGATGATGACTGTGACCACTATGCTGGTGAATTCTACATCCTTG
GTGATGAAGCTGAAATGCCATGGGCTATGATCGTCTTACAAATATGAGATCACTGAGCAGCTCA
ATGCCCTGGATCTACACATCGGAGATAGATTCTCATCAGATACGAAGCGTTGATCTCATGGTA
CAAGTCTGGAAGCAACATCTCCCCAACCTCTGTACATGACGAAGGGCAG

INTRON 2C/2D (SEQ ID NO:130)

GTGAGAACATTGATAATAGTCAAATGAAGTATATCCGATTCAAGCTGTCATAAGATGAGATA
CATAAATCACAAATGTTGTATTAGATATCTCTTAAATTAAATGCCGCTTTATCAATATTCGAGCA
ATCCTTCAGCAACATACACCAGCAAATGTTCATCAACAGACTATATTAAATATTTAAAAAT
CCTTCTCTGTTATAAATACCTAAAGTATCGAATTCTGAATGCGTCTCTGCAGCATATA
GTTAAGTTGTTGTTCTGTCAAG

DOMAIN 2D

GTCACCATTAGGCTGACGAGTACGACGAAGTTGTAAGTGCAGGCCACATCAGAAAGAATTAA
AAGATCTGTCAAAGGGAGAAGTAGAGAGAGCCTAAGGTCTGCCTCCTGCAACTTCAGAACGACGGAG
TCTATGAGAATATTGCCAAATTCCACGGCAAGCCTGGGTGATGATAACGGTCGAAGGTTG
CCTGTTGTCATGGAATGCCACCTTCCCCAGTGGCACAGACTCTATGTCCTCCAGGTGGAGA
ATGCTTGCTGGAGAGAGGATCTGCCGTCTGTGCCATACTGGGACTGGACTGAAACATTACAG
AGCTGCCATCTTGATTGCTGAGGCTACCTATTCAATTCCGTCACAAACGTTGACCTAATC
CTTCTTCAGAGGAAAATCAGTTGAGAATGCTGTTACAACACGTGATCCCAGCCTGAGCTGT
ACGTTAACAGGTACTACTACCAAAACGTATGTTGGCTTGAACAGGACAACACTGCGACTTCG
AGATACAGTTGAGATGGTCACAATGTTCTCATGCTGGCTTGGAGAGCTACTTATTCTA
TTTCTCTCTGATTATTCTGCATTGACCTGTGTTTCCCTCACCATGCGAACACAGATAGAT
TGTGGCCATCTGGCAGGAGCTGCAGAGGTACAGGAAGAACATACAATGAAGCGGATTGTGCCA
TTAACCTAATGCGAAACCTCTACATCCCTCGACAAACAGTGTCAATCATGATCCTGTAACCT
TTAAATACTCAAAACCCACTGATGGCTTGACTACCAGAACACTTGGATACAAGTATGACAACC
TTGAGTTCAATCATTCAGTATTCCAGGCTTGAAGAAATCATTGTTAGACAACGTCAAGATC
GTGTGTTGCAGGATTCTCCTTCACAACATTGGACATCCGCAACTGTTGAGATATTGTCGTTG
TCCCTACCACCGCGGTGAGCAAAACTGTGAAAACAAAGCCGAACATTGCGTACTCGGAGGAG
AAACAGAGATGGCGTTCATTTGACAGACTCTACAGGTTGACATCAGTGAACACTGAGGGACC
TCGGCATACAGCTGGACAGCCATGACTTGACCTCAGCATCAAGATTCAAGGAGTAAATGGATCCT
ACCTTGATCCACACATCCTGCCAGAGCCATCCTGATTTGCGCTGGTTCAA

INTRON 2D/2E (SEQ ID NO:131)

GTAAGAAAGTTCACTGCTAAATCTTTTATGATAGAGGGTAGAGAAAGTGGAGACAATGTGAC
AATATATTGAATAAAGTTGTTAAAATTAACTCTCATAGTTCAATTATGCTGAAGCTGTAG
CCATCTATAACTGTGTAACATGAAATGTTAAGACATTAACCTAAATACCTCAGCTGATAACAAAAC
AATGTTAATACATACGTCAATGTAACATTCTTATCTTAGGTTAGCATAAACACTTCAGAGA
TACAGTGACGAAAACCTCTATTAAATATTCTGAG

DOMAIN 2E

GTTCTTCCTGCGCCTGATGGCATTGAGATGACATCCTGTGAGAAAAGAAGTGAACAGCCTGA
CAACCAGGGAGACTGCATCTGATCCATGCTCTGAAAAGTATGCAGGAAGACCATTCACCTGATG
GGTCCAAGCCATTGCCTCTTCCATGCCCTGCCACCACTCTGCCCTCACCATGCAACTCACC
GTTATGCTTGCTGTCCACGGCATGGCTACATTCCCCAGTGGCACAGACTGTACACTGTACAGT

Fig. 6e

TCCAGGATGCACTGAGGAGACATGGAGCTGCAGTAGGTGTACCGTATTGGATTGGCTGCGACCGC
 AGTCTCACCTACCAGAGCTTGTCAACCATGGAGACATACCATGATATTGGAGTAACAGAGATTCC
 CCAATCCTTCTACCAAGCCAATTGAGTTGAAGGAGAAAACATTACAACAGAGAGAAGTCA
 TTGCAGACAAACTTTGTCAAAGGTGGACACGTTTGATAACTGGTCTCAAACAAGCCATCC
 TAGCGCTTGAGCAGGAAAACACTGTGACTTGAGATTGAGTTGAAATTCTTCACAACGGCGTT
 ACACGTGGTCGGAGGCAGTCGTACCCACTCTATCGGACATCTCCATTACGCATCCTACGACCC
 TTTCTACCTCCACCATTCCCAGACAGACCGTATTGGCAATCTGGCAAGAACACTCCAGGAACAGA
 GAGGGCTCTCAGGTGATGAGGCTCACTGTGCTCTCGAGCAAATGAGAGAACATTGAAGCCTTCA
 GCTTCGGCGCTCCTTATAACTGAATCAGCTAACACAGGATTCTCCGACCCGAGGACACCTTC
 ACTACAGGAAGTTGGTTATGAATATGACAATTAGAATTCTCTAGGAATGTCAGTTGCTGAACTGG
 ATCAATACATTATTGAACATCAAGAAAATGATAGAGTATTGCTGGGCTCTGTTGAGTGGATTG
 GAGGTTCCGCATCAGTTAATTCCAGGTTGAGAGCTGATTCCACATGTCAGGATGCTGGTACT
 TCACCGTTCTGGTGGCAGTGCTGAGATGGCGTGGCATTGACAGGCTATACAAATATGACATTA
 CTGAAACTCTGGAGAAAATGCACCTCGATATGATGACTTCACAATCTCTGTCAGTCTGACCG
 CCAACAAACGGAACGTGCTTGAGCAGCAGTCTAACCCAACACCGAGTGTATTCAGCGGGGAC
 ATC

INTRON 2E/2F-1 (SEQ ID NO:132)

GTAAGTAGTAAACTGCTCAGATTGTTTCATAATTACTCCACTATTAAGTAAAAAGTACTAGTAAT
 TCAATAGTACTGTCACAGAGAAATGTAACACAATAGACCACAGACTCCATTGTTAAACGCC
 GGCTTGGTAAGTCTGAGATTGGTACTGATGGAAGCTAAAATATTTGACAG

DOMAIN 2F-1 (1st part of domain f)

GTGACATAAAATACCAAGAGCATGTCAGCGAACCGTGGCGCGTGAGCTGAGCGATCTGCTGCGA
 GGGACCCGTCTAGTCTCAAGTCTGCTCTCGAGACCTACAGGAGGATGATGGCCCCAACGGATACC
 AGGCTTGCAGCCTCCATGGCTACCAGCAGGCTGCCATGATGCCAGGGAAATGAG

INTRON 2F-1/2F-2 (SEQ ID NO:133)

GTATATTAAAGTATTTATCTTACGCATGACCCCTGACCCATTATTTTTTAATCCTCGGATT
 TGTTTAATCCTGTTACCAGCGAACGGTCCGGGTTAGAATTGATCTTCAGTCACATTCTTGTGTA
 GGACTAACGAGTTGCTGGCTTGTTACTCGGTTGACACGTCAACGGATCCAATTGCAATTAG
 ATCGATGCTCATGCTGTTGATCCCTGGATTGCCTGGACTCCACATACCGCCGCATATTGC
 TGGTATATTGTCGAATGCGACGCTAACAGCAAGCCAACAAACTGAGACCTGGTGGTACAT
 GTCAGTTCTCTATTGCTGGGTTCAAACATAGCCATCAGTTGAAATATTCATACATAGAAGAAT
 ACCTCTGAATATGATGATGAAACATTACTAGACTTGCTGTGAGCCCCAGGCAAAATGCACTGT
 AAAAATACACTGACAGAGGATTAGGCATTCTGGGAGTACTGTATAGTTAGTGCATACATATTAG
 CGTCCCTCACTAAAACGAATCTCTGAATGCTATCAATTAAAGATCATGATGCTTGATTGTGCT
 ACTGTATTTAAAATGGTGTAAAGATTGCAATTACAATATACACAAACACGTTCTGCATCTCGG
 AGAATGCAATCTTCGTTGACCGTCTGTTCATATTTTATGCATGTAGTTGCACTACTTAG
 CGTCCAATAAATCATTACACAAAATCACACAAACAAACGATTAGGAATGTA
 CTGACTGTAGCTGCAA
 CGAATATACTGATCCTTCTTGTCCAG

DOMAIN 2F-2 (2nd part of domain f)

ATCGCATGTTGCATTACGGTATGCCGACCTTCCCCAGTGGCACAGACTGTACACCCTGCAGTTG
 GAGATGGCTCTGAGGAGACATGGATCATCTGTCGCCATCCCCTACTGGACTGGACAAAGCCTATC

Fig. 6f

TCCGAACCTCCCTCGCTTCAACCAGCCCTGAGTATTATGACCCATGGCATGATGCTGTGGTAAAC
 AACCCATTCTCAAAGGTTGTCAAATTGCAAATACCTACACAGTAAGAGACCCACAGGAGATG
 CTGTTCCAGCTTGTGAACATGGAGAGTCACATCCTCTATGAGCAAACACTCTCTGCTCTAGAGCAA
 ACCGACTACTGTGATTGAGGTACAGTTGAGGTCCATAACGTGATCCACTACCTTGTG
 GGACGTCAGACCTACGCATTGTCTCTGCATTATGCATCCTACGACCCATTCTTACAC
 CATTCTTGTGGATAAGATGTGGTAGTATGGCAAGCTCTCAAAAGAGGAGGAAACTCCATAC
 AAGCGAGCTGACTGTGCTGCAACCTAATGACTAAACCAATGAGGCCATTGACTCCGATATGAAT
 CAGAACCCATTACAACAGATGCACGCAGTCCCAACACACTCTATGACTACGAGACACTGTACTAC
 AGCTACGATAATCTGAAATAGGTGGCAGGAATCTGACCAGCTCAGGCTGAATTGACAGAAC
 AGAACGCCACGATCGCGTTTGCTGGATTCTGCTCGTGGAACTCGAACCTCTGCTGATGTCAGG
 TTTGGATTGAGAAATGAAAATGACTGCCACAGGGTGGAAATAATTTCATCTTAGGTGGAGCC
 AAGGAAATGCCATGGTCATTGACAGAAACTTCAAGTTGATATCACCCATGTAACGAGAAC
 GGCATTAGCCCAGAGGACGTGTTGATGCTGAGGAGCCATTATATCAAGGTTGAGATCCATGCT
 GTTAACAAGACCATGATACCATCGTCTGTGATCCCAGCCCCAACTATCATCTATTCTCCTGGGAA
 G

INTRON 2F-2/2G-1 (SEQ ID NO:134)

GTGAGAGAACCAAGTAATAGCTACTGTCTACAAAGAACATGTGTTCATTAAAGACCTGACTGTAGGCC
 GATGGCTGCTGTCATCTCCCGCCCTCCTCTGTTCTCCTCCGAAGGGGTAGCTTCAGGTT
 CTCTTGCCAATATGCCAACGACCTCCTGAGCAGGCAGTATATACGTAAGGGAAAGCAAGTATG
 GACCATCGCGGGCATGTAGAGATAACATGATCAGCTGCTGCTGTTCCACTCCTGTCAGACAATG
 AGATAAACATGAATACAGTATTACTCAGCAGCGTCCAATTTCACCCCTCGTATTAAAAAA
 AGGAATTTTAATATATTTCCTCTGTTGAAATATTTAGTAACGTTAATCGATATAGAGTGG
 AGTAGTGACGTTTATTCGGTCATTCTCGAAACAAAATATAATAGTCCACTGAACCTCTTAA
 ATTGTTTACAACCTCACTGCCACAGACGTAATCCCTCACGTTATTGAGCTGACAACGTGT
 TGAATTGAGTGTGTTCCGAATTCTAAATAAGCATGTATATATTACGTCATGCAAGTAATATAT
 GTTTAACTGATGACGTCACTGGTGACCACTGATTAGTCCTTGTCTGATTCAGTTCTGTT
 GTCACGGGACGGTGGGAAGCCAGGTTCCCTGTCAGCTGAATATCCGTCGAATCCCCAC
 ATGGGTACAAAGTGTGATGCCTATTCTGGTGTCCCCACCGTGAATTGCTGGAATAAGTGGCTT
 AATACCATATACACTCACTCTATTGTCACACTACTGCCACCGCTCACACCTCTGATGCTTCTGTT
 CTATCCAG

DOMAIN 2G-1 (1st part of domain g)

GTCGCCTGCTGACAGTCACACTCAGCCAACATTGCTGGCTCTGGGGTGAGGAAGGACGTCACGA
 CCCTCACTGTCGAGACCGAGAACCTAACAGACAGGCTCTCAAGGTGTCATCGATGATACTGGTC
 CCAATGGTTACCAAGCAATAGCATCCTCACGGAAGTCCTCCAATGTGCGAGATGAACGGCCGCA
 AGGTTGCCTGTTGCTCACG

INTRON 2G-1/2G-2 (SEQ ID NO:135)

GTAATTAATGGATGTGAAGTCAATGTCCGAGGGTATAATAAGGATTAAATAACTTCAGTCGTGAA
 TACTGTATGACATGTGATTGGATGGTAGGTATTACAGGTTATAAGGCCAGTGTGTTGGGAC
 GGTTACTTCCCTGCACTAGTAATAAGCATTGTATTAGCTAGCTTATCATATAACCTTAGTT
 ATGGTTGTGGCAATTGAAATCGAAATTCTTCAAGGTATCGCACTCGTGTGTTAGAA
 TAGTTACTATGCTGCATTGAGAATAACACTATAGTAATAAGCATATCATACTAGTAAGAATAACAC
 TATAGTAATAAAAGTATATCATACAGTAAGAATGTCATTGTATGATAAATAGGTTATCACACTCGT
 TGTTTACAAGTGGTACTATCCAGGAATAACCACTATGTATTACATGTATATTGGCAGTGTAA
 TAGTAGCATTGTATATTAAATCAGTATATCGTCAAAACACCAGGATATATGGGTATACAGT

Fig. 6g

GGGCAGTGTAAAGTAGCAACATTGTATATTAAATCAGTATATCGTACTTCAAAACACCAGGATTATG
 GGGTATACTACAGTGGCAGTGTAAAGTAGTAGCATTGTATATTAAATCAGTATATCGTACTTCAAAACA
 CCAGGATATAATTCAAGTATATCGTCTCAAAACACCAGGATATAATTCAAGTATATCGTCTCAA
 AACACCAGGATATGGGATATACTAGTGCAGGTTGCATAAACCTCCACCCTTACAG

DOMAIN 2G-2 (2nd part of domain g)

GTATGGCCTCCTCCCACACTGGCACAGACTGTATGTGAAGCAGATGGAAGACGCCCTGGCTGACC
 ACGGATCACATATCGGCATCCCTACTGGACTGGACAACGCCTCACAGAGTTACCCGCCCTTG
 TCACAGACTCCGAGAACAAATCCCTCCATGAG

INTRON 2G-2/2G-3 (SEQ ID NO:136)

GTCAGTTAGTCTCCTGTCTGAGCTAACGATACCAATTCCATTTCGAGAACACCACGATGACGAG
 AAAACAAGCAATATAGATATAGATGCAGTATAGATCAAGTTAATGAATTCTTGCTATATGTTGC
 TTGTAATAAAACTTAAGAAAACGAGAGCATGCACACAAATGAAACAAACAATTATGTGTTGATAG
 GAATATGATATATGTATTGGGGCTGACGTGAGCAGGGTTGAAGGGACAGTTACATTGTCAGTA
 ACACGGAGTATTCTTGATCCACAATATAGTTCAATTGTGTTCAGCAGTTACAACAACTACATT
 ATATCATACATTACGTCGAACATGCTCTTGTCTCTGCCAG

DOMAIN G-3 (3rd part of domain g)

GGTCGCATTGATCATCTCGGTGTAAACCACGTACGTTCCCCCAGAGACATGCTTTAACGACCCA
 GAGCAAGGATCAGAGTCGTTCTTCTATAGACAAGTCCTGGCTTGGAGCAGACTGACTACTGC
 CAGTCGAAGTCCAGTTGAGCTGACCCACAACGCCATTCACTCCTGGACAGGGACGTAGCCCT
 TACGGAATGTCGACCCCTGAGTTACAGCCTACGATCCTCTCTGGCTTACCACTCCAACACC
 GACAGAATCTGGCTGTCTGGCAAGCACTGCAGAAATACCGAGGACTCCATACAACGAAGCACAC
 TGTGAAATCCAGGTTCTGAAACAGCCATTCAACGATGACATCAACCACAATCCAATC
 ACCAAGACTAATGCCAGGCCTATCGATTGATTGAGAGGTTAACTATCAGTATGACACC
 CTTAGCTTCCATGGTAAGAGCATCCCTGAACCTGAATGACCTGCTCGAGGAAAGAAAAAGAGAAGAG
 AGAACATTGCTGCCCTCTTCTCGTGGAAATCGGTTGCAGTGCTGATGTCGTTGACATCTGC
 CGCCCCAATGGTACTGTGCTTGCAGGAACCTTGCTGTGCTGGAGGGAGCTAGAAATGCC
 TGGCCTTCGACAGACTGTTCCGCTATGACATCACCAGAGTCATGAATCAGCTCCATCTCCAGTAT
 GATTCAAGATTCAGTTCAAGGTGAAGCTTGTGCAACCAATGGCACTGAGCTTCATCAGACCTC
 CTCAAGTCACCAACAATTGAACATGAACATTGGAG

INTRON 2G-3/2H (SEQ ID NO:137)

GTATGTTATCTTATTCAAATGTGTAACTCAGACTGGAGACGTTTCAATTAACTTGGTCAGC
 ATTAGTTGATGATTTGGTGCATATTGACGACAAGGAGTTAACGATTAACACGTTAACACATCT
 TTAATCTGATATGAGAAGGAAATAATTGATCCAGTATTGATGATTGAAGTTAGATTAAACAGTGA
 AGATATAACCAAGTTGATAATCGTATAAAACAGTAGCAGAATTGTATCGTAAAACAAATGTGGG
 AAGGCGAACGCCAACGAGATTAGATTACGATCGTGTGCTAGAATAATTACAATAACCCAGACG
 TCGGAAATGTGGTGTCTATGGCAATAGTTACGATTAATTGCTAACATGCACGATTACCTATT
 AG

DOMAIN 2H

CCCACAGAGGACCAGTTGAAGAACAGAACGACTCACCAAAACTGACGGCAATGCACACTTCC
 ATCGTAAGGAAGTTGATTGCTGTCCCTGGATGAAGCAAACAACTTGAAGAATGCCCTTACAAGC

Fig. 6h

TACAGAACGACCACAGTCTAACAGGATACGAAGCAATCTGGTTACCATGGATACCCGAATCTGT
 GTCCGGAAGAAGGCATGACAAATACCCCTGCTGCGTCCACGGAATGCCATCTCCCCACTGGC
 ACAGACTCTTGACCATCCAACGGAAAGAGCTCTCGAGCACAAATGGTCAGTGCCTGGTGTCTT
 ACTGGGACTGGACCAAGGACCTGTCGTACTGCCGGCTTCTCCGACTCCAGCAACAACAATC
 CCTACTTCAAGTACCAACATCGCAGGTGTTGTCACGACACCGTCAGAGAGCCAATAGTCTTATAT
 ATAACCAGCCCCAAATCCATGGTTATGATTATCTTATTACCTAGCATTGACCACGCTTGAAGAAA
 ACAATTACTGTGACTTTGAGGTTCACTATGAGATCCTCCACAACGCCGTCCACTCCTGGCTGGAG
 GATCCCAGAAGTATTCCATGTCTACCCCTGGAGTATTGGCCTTGACCCGTCTTATGATCCTTC
 ACTCGGGTCTAGACAGACTTGGATCATCTGGCAAGAACCTCAGAAGATCAGGAGAAAGCCCTACA
 ACTTCGCTAAATGTGCTTATCATATGATGGAAGAGCCACTGGCGCCCTCAGCTATCCATCTATCA
 ACCAGGACGAGTTCACCGTGCACCTCAAGCCTTCTACAGTTTGACAGCCATAAGTCGGCT
 ACCATTACGATAACCTGAATGTTAGAGGTACAGCATCCAAGAACTCAACACAATCATCAATGACT
 TGAGAAACACAGACAGAATCTACGAGGATTGTTGTCAGGCATCGGTACGTCTGCTAGTGTCA
 AGATCTATCTCGAACAGATGACAATGACGAAGAAGTTGAACTTCACTGTCCTGGGAGGAGAGA
 GGGAAATGCCATGGCCTACGAGCAGTTCAAGTATGACATCACAGAGGTTGCAGATAGACTTA
 AACTAAGTTATGGGACACCTTAACCTCCGACTAGAGATCACATCCTACGATGGATCGGTGGTAA
 ACAAGAGCCTACCAATCCTTCATCATCTACAGACCTGCCAATCATGACTACGATGTTCTGTTA
 TCCCAGTAGGAAGAACCTTCACATCCCTCCAAAGTTGTCGTCAAGAGAGGCACCGCAGTCA
 TCCACCCAGTCGATGATTCACTACGAGACCAGTTGTTGATCTTGGAAAGCTACACTGCACCTTCA
 ACTGTGTGGTACCACCGTTCACATACCGCGATTGAACTGAACCACGTCTATTGTCAAGCCTG
 GTGACTACTATGTTACCGGACCAACGAGAGACCTTGCCAGAATGCAGATGTCAGGATTCAATATCC
 ATGTTGAGGATGAGTAA

3' UTR

CGAACACAG

INTRON 3'UTR (SEQ ID NO:138)

GTGAGATAAGAAACCCCTAACAGTAATACGACACCAATTACAGCTAAACATGATTGCCATCG
 ATGTTTCATGTGAGTATACGCTTTCACTACATAATTGTTTCAAATCAAGTTAGCA
 AATGAATCTATCACTGGAAAATAGGGTAGGGTAGCCAAGTGGTAAAGCGGTCACTGATCACGCCA
 AAGACGAGTGTCTAACCTGCATGGTACAAAAGTGAAGACCATTGCTGGTGTCTACCGCCGTAAT
 ATTGTTTAGTATTGCTAAAACCTATACTCACCCATGCGCTGAAAGTGGATAATAATCATAT
 TTCAACAAAAGCACAAACCATTCATGAAAGCCTTGTACCTGAAAGACGCAAGAG
 AACAAATAGTCCTAACATTATTCAGACATTGAAATGTCCTGCACGTGAAACCATAATCCTT
 TGAAATTTACGACTGCATCGTATACAATTATGATAAAATTAAAACCTTATTCAG

3' UTR

GTTCTTGGTCTCCACATATTCACACATCAGCACAAACGGTTCGAAGGACATTGGCGTTCTTCT
 CTGGCAATGCATTCAATACACATTGAAAATGACTTCAGCATATCAGTGTGCTTCGAACGTGTTC
 CGGAAGTACTCAAATGTGCTATGACTGAATTATTGTACATACATAACTTATTGATGTTCAATAAAAT
 AAATGTTGAAACG

Fig. 7aPrimary structure of the HtH2 proteinDOMAIN A (SEQ ID NO:156)

GLPYWDWTQHLTQLPDLVSDPLFVDPEGGKAHDNAWYRGNIKFENKKTARAVDDRLFEKVGPGENT
 RLFEGLDALEQDEFNCNFEIQFELAHNAIHVLVGGRHTYSMSHLEYTSYDPLFLHHSNTDRIFAI
 WQRLQVLRGKDNPNTADCAHNLIHEPMEPFRRDSNPLDLTRENSKPIDSFDYAHLYQYDDLTLNGM
 TPEELNSYLHERSGKEGVFASFRLSGFSSANVVVYACRPAHDEMAVDQCDKAGDFVLGGPTEMP
 WRFYRAFHFDVTDSIDNIDKDRHGHYYVKAEELFSVNGSALPNLQPPTISHRPARGHVDEAPAPS
 SDAHLAVRKDINHLTREEVYELRRAMERFQADTSVDGYQATVEYHGLPARCPFPEATNRFACCIHG
 MATFPHW

DOMAIN B

HRLFVTQVEDALIRRGSPIGVPYWDWTQPMALPGLADNATYRDPISGDSRHNPFHDVEVAFENGR
 TERHPDSRLFEQPLFGKHTRLFDSIVYAFEQEDFCDFEVQFEMTHNNIHAWIGGGKYSMSSLHYT
 AFDPISYLNHSNTDRLWAIWQALQIRRNKPYKAHCWSEERQPLKPFSSPLNNNEKYENSVPT
 NVYDYEGVLGYTYDDLNFGGMDLGQLEEYIQRQRQRDRTFAGFFLSHIGTSANVEIIIDHGTLLHTS
 VGTFAVLGGEKEMKWGFDRLYKYEITDELRQLNLADDGFSISVKVTDVDGSELSSELIPSAIIIF
 ERSH

DOMAIN C

IDHQDPHQDTIIRKNVDNLTPPEEINSLRRAMADLQSDKTAGGFQQIAAFHGEPKWCPSPDAEKKFS
 CCVHGMASFPHWHRLLTQGENALRKHGCLGALPYWDWTRPLSHLPDLVSQQNYTDAISTVEARNP
 WYSGHIDTVGVDTRSRVQELYEAPGFGHYTGVAQVLLALEQDDFCDFEVQFEIAHNFIAHALVGG
 SEPYGMASLRYTTYDPIFYLHHSNTDRLWAIWQALQKYRGKPYNANCIAIASMRKPLQPFGLTDEI
 NPDDETRQHAVPFSVFDYKNNFNYEYDTLDFNGLSISQLDRELSRRKSHDRVFAGFLLHGIQQSAL
 VKFFVCKSDDDCDHYAGEFYILGDEAEMPWGYDRLYKYEITEQLNALDLHIGDRFFIRYEAFDLHG
 TSLGSNIFPKPSVIHDEGA

DOMAIN D

GHHQADEVVTAASHIRKNLKDL SKGEVESL RSAFLQLQNDGVYENIAKFHGKPGLCDDNGRKV
 ACCVHGMPTFPQWHRLYVLQVENALLERGSAVSPYWDWTETFTELPSLIAEATYFNSRQQTDFPN
 PFFRGKISFENAVTTRDPQPELYVNRYYQNVMLAFEQDNYCDFEIQFEMVHNVLHAWLGGRTAYS
 ISSLDYSAFDPVFFLHHANTDRLWAIWQELQRYRKPKYNEADCACINLMRKPLHPFDNSDLNHDPT
 FKYSKPTDGF DYQNNFGYKYDNLEFNHSIPRLEEIIRIRQRQDRV FAGFLLHNIGTSATVEIFVC
 VPTTSGEQNCENKAGTFAVLGGETEMAFHFDRLYRFDISETLRDLGIQLDSHDFDLSIKIQGVNGS
 YLDPHILPEPSLIFVPGSS

DOMAIN E

SFLRPDGHSDDILVRKEVNSLTTRETASLIHALKSMQEDHS PDGFQAIASFHALPPLCPSPSATHR
 YACCVHGMATFPQWHRLYTVQFQDALRRHGAAGVGPYWDWLRPQSHLPELVTMETYHDIWSNRDFP
 NPFYQANIEFEGENITTEREVIADKLKVKGHVFDNWFFKQAILALEQENYCDFEIQFEILHNGVH
 TWVGGSRTHSIGHLYASYDPLFYLHHSQTDRIWAIWQELQEQRGLSGDEAHCALEQMREPLKPF
 FGAPYNLNQLTQDFS RPE DTFDYRKFGYEDNLEFLGMSVAELDQYIEHQENDRVFAGFLLSGFG
 GSASVNFQVCRADSTCQDAGYFTVLGGS AEMA WFDR LYKYDITETLEKMHLRYDDDFTISVSLTA
 NNGTVLSSLIPTPSVIFQRGH

Fig. 7bDOMAIN F

RDINTKSM SANVRRELSDLSARDPSSLKSALRDLQEDDGPNQYQALAAFHGLPAGCHDSQGNEIA
 CCIHGMPTFPQWHRLYTLQLEMALRRHGSSVAIPYWDWTKPISEPLSLFTSPEYYDPWHDAVVNNP
 FSKGFVFKFANTYTVRDPQEMLFQLCEHGESILYEQTLLALEQTDYCDFEVQFEVLHNVIHYLVGGR
 QTYALSSLHYASYDPFFFIIHHSFVDKMWVVQALQKRRKLPYKRADCAVNLMTPKPMRPFDSMDMNQN
 PFTKMHAVPNTLYDYETLYSYDNLEIGGRNLQAEIDRSRSHDRVFAGFLLRGIGTSADVRFW
 ICRNENDCHRGGIIFILGGAKEMPWSFDRNFKF DITHVLEKAGISPEDVFDAEEPFYIKVEIHVN
 KTMIPSSVIPAPTIIYSPGE

DOMAIN G

GRAADSAHSANIAGSGV RKDV TTLTVSETENLRQALQGV IDDTGPNGYQAIASFHGSPPMCEMNGR
 KVACCAHGMASFPWHRLYVKQM EDALADHGSHIGIPYWDWTTAFTELPALVTDSENNPFHEGRID
 HLGVTTSRSPRDMLFNDPEQGSEFFYRQVLLALEQTDYCQFEVQFELTHNAIHSWTGGRSPYGMS
 TLEFTAYDPLFWLHHNSNTDRIWAVWQALQKYRGLPYNEAHCEIQVLKQPLRPFNDDINHNPIKT
 ARPIDSF DYERFNYQYDTLSFHGKSIPELNDLLEERKREERTFAAFLLRGIGCSADVFDICRPN
 DCVFAGTFAVLGGELEMPWSFDRLFRYDITRVMNQLHLQYDSDFSFRVKLVATNGTELSSDLLKSP
 TIEHEL

DOMAIN H

GAHRGPVEETEVTHQNTDGNAHFHRKEVDSL SLD EANNLKNALYKLQNDHSLTG YEAI SGYHGYPN
 LCPEEGDDKYPCCVHGMAIFPHWHRLLTIQLERALEHNGALLGVPYWDWTKDLSSLPAFFSDSSNN
 NPYFKYHIAGVGHD TVREPTSLIYNQPQIHGYDYL YYLA TTLEENNYCDFEVQYEILHNAVHSWL
 GGSQKYSMSTLEYSAFDPVFMILHSGLDRLWI IWQELQKIRRKPYNFAKCAYHMMEPLAPFSYPS
 INQDEFTRANSKPSTVFD SHKFGYHYDNLNVRGHSIQELNTIINDLRNTDRIYAGFVLSGIGTSAS
 VKIYLRTDDNDEEVGTFTVLGGEREMPWAYERVFKYDITEVADRLKLSYGDTFNFRLEITSYDGSV
 VNKSLPNPFI IYR PANHDYDVLVI PVGRNLHIPPKV VVKGTRIEFHPVDDS VTRPVVDLGSYTAL
 FNCV VPPFTYRGFELNHV YSVKPGDYYVTGPT RDLCQNADV RIHIHVEDE

Fig 8aGenomic sequence of the KLH1 gene

DOMAIN 1B

GGCCTACCGTACTGGACTGGACTGAACCATGACACACATTCCGGGTCTGGCAGGAAACAAA
 TATGTGGATTCTCATGGTCATCCCACACAAATCCCTTTCATAGTTCACTGATTGCATTGAAGAA
 AATGCTCCCCACACCAAAGACAAATAGATCAAAGACTCTTAAACCCGCTACCTTGACACCAC
 ACAGACCTGTTCAACCAGATTGTATGCCTTGACAAGAAGATTACTGTGACTTGAAGTCCAA
 TTTGAGATTACCCATAACACGATTCACTGGACAGGAGGAAGCGAACATTCTCAATGTCGTCC
 CTACATTACACAGCTTCGATCCTTGTACTTCACCATTCTAACGTTGATCGTCTTGGGCC
 GTTGGCAAGCCTACAGATGAGACGGCATAAACCTACAGGGCCACTGCGCCATATCTCTGGAA
 CATATGCATCTGAAACCATTGCCCTTCACTCTCCCTTAACAATAACGAAAGACTCATGCCAAT
 GCCATGCCAAACAAGATCTACGACTATGAAAATGTCCTCATTACACATACGAAGAGATTAAACATT
 GGAGGCATCTCTGGAAAACATAGAAAAGATGATCCACGAAAACAGCAAGAAGACAGAAATAT
 GCCGGTTTCTCCTGGCTGGCATACGTACTTCAGCAAATGTTGATATCTTCACTAAACTACCGAT
 TCCGTGCAACATAAGGCTGGAACATTGCACTGCTCGTGGAAAGCAAGGAAATGAAGTGGGATT
 GATCGCGTTTCAAGTTGACATCACGCACGTTGAAAGATCTGATCTCACTGCTGATGGCGAT
 TTCGAAGTTACTGTTGACATCACTGAAGTCGATGGAACATAACTGCATCCAGTCTTATTCCACAT
 GCTTCTGTCATTCGTGAGCATGCACGTGGTAAGCTGAATAGAG

INTRON 1B/1C (SEQ ID NO:139)

GTTTGTAATAATTATGTAGAATTCTTACCTCAGAATAAGATGAGGTACATGGGTTTGCAAAA
 CTATTACGTTCGAATTAATATTAATAATACCGGACCCCTCCACTGGTACATATTATCTTATAACG
 ATAATAGCGATGATGATGATGATGATGATGATGATAATGATGATGCCGGTATTG
 CACGTAATCCAGCCGACTTAGATGACACCCTAACGGTGCAGAAAGTATAACAATTAGATTGCGTT
 GCATCTGTGTATGCGTGTGCTTAACCAAAAGTCAAAATAAAAGTCAAACCTTAGTTATTCA
 TTGATAGAGCCTTACGATAAGAACAAATGTAATAAAATTAGAACATAACTGAAACCTCCGAAAGAA
 GGCCTGTTGTCAAGAGAGGTATCGACATGATTGACTTATAAACCTGTGCTCTATATTGGAAC
 TGTCCACTTCTTGTGTACTGTAATCACATCGACTATGGCTGCAAGACGTGTACGAGTAC
 ACTATATACTTACCTAATGACCAACCACAAGGCTGGCTTGTAAATTGTTATTACAGAAATA
 AACACAGAATTCCAGCATTTGGCTGGTGTATTAGCAAAACACCGATATGACACTCATGTTTATT
 ACATTTTTTCAG

DOMAIN 1C

TTAAATTGACAAAGTCCAAGGAGTCGTTATTGAAAAAATGTAGACCGTTGAGCCCCGAGG
 AGATGAATGAACCTCGTAAAGCCCTAGCCTACTGAAAGAGGGACAAAGTGCCGGTGGATTCAGC
 AGCTTGGTGCATCCATGGGAGCCAAATGGTGTCTAGTCCCAGCATCTAAAAAATTGCGCT
 GCTGTGTTCACGGCATGTCGTGTTCCCTACTGGCATCGACTGTTGACGGTCAGAGTGAATG
 CTTTGAGACGACATGGCTACGATGGAGCTTGCCGTACTGGGATTGGACCTCTCCTCTTAATCACC
 TTCCCGAAGTGGCAGATCATGAGAAGTACGTCGACCCCTGAAGATGGGTAGAGAACGATAACCTT
 GGTCGATGGCATATAGATACAGTCGACAAAACAACAAGAAGTGTTCAGAATAAACTCTCG
 AACAGCCTGAGTTGGTATTACAGCATGCCAAACAAGTACTGCTAGCGTTGGAACAGGACA
 ATTCTGTGACTTGAATCCAATATGAGATTGCCATAACTACATCCATGCACTGTAGGAGGCG
 CTCAGCCTTATGGTATGGCATCGCTCGTACACTGCTTTGATCCACTATTCTACTTGCATCACT
 CTAATACAGATCGTATATGGCAATATGGCAGGCTTACAGAAGTACAGAGGAAACCGTACAACG
 TTGCTAATGTCGTGTTACATCGATGAGAGAACCTTGCAACCATTGGCCTCTGCCAATATCA
 ACACAGACCATGTAACCAAGGAGCATTCACTGCAACGTTTGATTACAAGACCAATTCA
 ATTATGAATATGACACTTGGATTAAACGGTCTCTCAATCTCTCAGTTGAATAAGCTCGAAG

Fig. 8b

CGATAAAAGAGCCAAGACAGGTTCTTGCAGGCTTCCTGTTATCTGGTTCAAGAAAATCATCTCTTG
 TTAAATTCAATATTGCACCGATAGCAGCAACTGTCACCCCGCTGGAGAGTTTACCTCTGGGTG
 ATGAAAACGAGATGCCATGGGCATACGATAGAGTCTCAAATATGACATAACCGAAAAACTCCACG
 ATCTAAAGCTGCATGCAGAAGACCACCTCTACATTGACTATGAAGTATTGACCTAAACCAGCAA
 GCCTGGAAAAGATTGTTCAAGCAGCCTCAGTCATTGATGAACCAAGAATAG

INTRON 1C/1D (SEQ ID NO:140)

GTACTTGTATATGTTCGAATATTGCCGATACTTCAATATATATACTTTATCAAAGTAATTGAT
 TAATCTGAAGTAATTTCTTCCAGTAGAGATTCACTGATACAACAAGAATTGCCCTGTTGTA
 TGTCACTTATTTCATCAAACGATTGAGCTGTCCATGCCACAATGGGTCTCTGTAAC
 TTCTCGTATGGGTATAGATTATAGACGTGGCAGACCTTACGTATAACTAATATTGTTGTAATG
 TCGTTTCAG

DOMAIN 1D

GTCACCATGAAGGCGAAGTATATCAAGCTGAAGTAACCTCTGCCAACCGTATTGAAAAAACATTG
 AAAATCTGAGCCTGGTGAACTCGAAAGTCTGAGAGCTGCCCTGGAAATTGAAAACGATGGAA
 CTTACGAATCAATAGCTAAATTCCATGGTAGCCCTGGTTGTGCCAGTTAAATGGTAACCCCATCT
 CTTGTTGTGTCCATGGCATGCCACTTCCCTCACTGGCACAGACTGTACGTGGTTGTGTTGAGA
 ATGCCCTCCTGAAAAAAGGATCATCTGTAGCTGTTCCCTATTGGACTGGACAAAACGAATCGAAC
 ATTACCTCACCTGATTCAGACGCCACTTACTACAATTCCAGGCAACATCACTATGAGACAAACC
 CATTCCATCATGGAAAATCACACACAGAGAATCACTACTAGGGATCCCAAGGACAGCCTCT
 TCCATTCAAGACTACTTTACGAGCAGGTCTTACGCCCTGGAGCAGGATAACTCTGTGATTTCG
 AGATTCAAGTGGAGATTACACAATGCATTGATCTTACTTGGTGGCAAAGGTAATATTCCA
 TGTCAAACCTTGATTACGCTGCTTGTGATCCTGTGTTCTCCTCATCAGCAACGACTGACAGAA
 TCTGGCAATCTGGCAAGACCTTCAGAGGTTCCGAAAACGCCATACCGAGAAGCGAATTGCGCTA
 TCCAATTGATGCACACGCCACTCCAGCCGTTGATAAGAGCGACAACAATGACGAGGCAACGAAA
 CGCATGCCACTCCACATGATGGTTTGAAATATCAAACAGCTTGGTTATGCTTACGATAATCTGG
 AACTGAATCACTACTCGATTCCCTCAGCTGATCACATGCTGCAAGAAAGAAAAGGATGACAGAG
 TATTGCTGGCTCCTCCTCACAATATTGGAACATCTGCCGATGCCATGTATTGTATGTCTCC
 CAACTGGGAAACACACGAAGGACTGCAGTCATGAGGCTGGTATGTTCTCCATCTAGGGCGTCAAA
 CGGAGATGTCCTTGTATTGACAGACTTACAAACTTGACATAACTAAAGCCTGAAAAAGAACG
 GTGTGCACCTGCAAGGGATTTCGATCTGAAATTGAGATTACGGCTGTGAATGGATCTCATCTAG
 ACAGTCATGTCATCCACTCTCCACTATACTGTTGAGGCCGGAACAG

INTRON 1D/1E (SEQ ID NO:141)

GTAACATTTCGACTGTAACCAACAACGCTGAGTCTATTGCAATTACGATAATAACAATTTC
 GAAATATATCTTATTAAAGCAAAGGTTCTAGAGACAAACAGCCGCTCTAATTATTTCGAA
 CTTACGCTTGAGTAAAGATCTGCAAATGCCAACCCCTACCTATACTATTAAAATAATGTTACAT
 TCGTATCTGAATGTTAATAAAATCACTTCATATTCTGTTGCAG

DOMAIN 1E

ATTCTGCCACACAGATGATGGACACACTGAACCAGTGTGATTGCAAAAGATATCACACAATTGG
 ACAAGCGTCAACAACACTGTCACTGGTGAAGGCCCTCGAGTCCATGAAAGCCGACCATTCTGATG
 GGTTCCAGGCAATCGCTTCTTCCATGCTCTCTCTTGTCCATCACCAGCTGCTCAAAGA
 GGTTGCGTGCTCGTCCATGGCATGGCAACGTTCCCACAATGGCACCGTCTGTACACAGTCCAAT
 TCCAAGATTCTCTCAGAAAACATGGTGCAGTCGTTGGACTTCCGTACTGGGACTGGACCCCTACCTC

Fig. 8c

GTTCTGAATTACCAGAGCTCCTGACCGTCTCAACTATTCATGACCCGGAGACAGGCAGAGATATAC
CAAATCCATTATTGGTTCTAAAATAGAGTTGAAGGGAGAAAACGTACATACTAAAGAGATATCA
ATAGGGATCGTCTCTTCCAGGGATCAACAAAAACATCATTAACGGTTATTGAGCAAGCACTGC
TTGCTCTGAACAAACCAACTACTGCGACTTCAGGTTCAAGTGCATTATGCTTCATGATCCAC
ATACCTGGGTGGAGGCAAGGAGCCCTATGGAATTGCCATCTGCATTATGCTTCATGATCCAC
TTTCTACATCCATCACTCCAAACTGATCGTATTGGCTATATGCAATCGTTGCAGCGTTCA
GAGGACTTCTGGATCTGAGGCTAAGTGTGTAAATCTCATGAAAACCTCTGAAGCCTTCA
GCTTGAGCACCATAATCTTAAATGATCACACGCATGATTCTCAAAGCCTGAAGATACTTCG
ACTACACAAAGTTGGATACATATGACACTCTGAAATTGCAGGGTGGTCAATTGGCATTG
ACCATATTGTCGTAACAGGCAGGAACATTCAAGGGCTTGCAGGATTCTGCTGAAGGATTG
GCACCTCTGCCACTGTCGATTCCAGGTCTGTCGACAGCAGGAGACTGTGAAGATGCAGGGTACT
TCACCGTGTGGAGGTAAAAAGAAATGCCATTGATCGCTTACAAGTACGACATAA
CAGAACCTTAGACAAGATGAACCTTCGACATGACGAAATCTCCAGATTGAAGTAACCATACAT
CCTACGATGGAACTGTACTCGATAGTGGCCTTATTCCCACACCGTCAATCATATGATCCTGCTC
ATC

INTRON 1E/1F (SEQ ID NO:142)

GTAAGTATAACACACATTATTCTCTCTGCTATATCAGATGAAGAGAACGTTGATCACTAACCTA
GTCTGTTGATTGTGGTTCTGTTGCTTCTGAACAGTAGGGTTGATTTAACTTCTCTGTTTCG
TCTGTACCAATGAAAGACTATGATGCTTGTGAAGATGCTTGTATGAGTCAGTCTGTTCTG
TAATGCTTGATCTTGCCATCAACATTGAAATTATGGTTCCCTAAATACTTACATA
TTACATTAAACGTCGCTGCTGTGATTGCATATTCTTCAAAAATAACTATATATTCCAG

DOMAIN 1F-1 (1st part of domain f)

ATGATATTAGTCGACCACCTGTCGCTCAACAAGGTTGTCATGATCTGAGTACACTGAGTGAGC
GAGATATTGGAAGCCTAAATATGCTTGAGCAGCTGCAGGAGATACTCAGCAGATGGTTTG
CTGCCATTGCATCCTCCATGGTCTGCCTGCCAAATGTAATGACAGGCCACAATAACGAG

INTRON 1F-1/1F-2 (SEQ ID NO:143)

GTAAATATAACAGTGAATCCGGATAAGTAAAATCCAGATAAGAAAAAAAAACATTCTGTGGTCCC
GGCATGTTCTCTCATCTATCATTATTGATAACGGATAAGTAAAATCGGCTGAGTAAAACAT
CCGGGTAAGTAAAATGATTTGAGGTCTCTCATGGATAAGTAAAGATAACAAAGTGTACATTCC
ATAAAACACTAATGATGCAACACAATACCAAGCGCACAGTGTGTTCACTACGTTGTTGTATTGT
ATTAACAATTAAACACTTAAGTGTGTTCCATGTCGTTATGTCTACGTTGTTATTCTGACTGCTGGAGGGTTC
CAACAAGCCCCGCAATTCCATGTCGTTATGTCTACGTTGTTATTCTGACTGCTGGAGGGTTC
GGAAAAAAATAAAAACGGTAAATATTATAAAAATTACCGGTGCCTGAAATTAGGTGTCG
GATTTCACTGTAGATGATTAATTCTCACTTGTAAACAAAAGGACCCAGTACCCCTATTGAC
GTACGTTATAAAATGTAATTATAAAAAGCCATTATCATGTTATACGTGATCTGNCTGCAATT
TNCTACCGCTTCTGATTAAAGCAATTCTCCCTATGAACATTAAACATAGCACTCCT
GCAAAAGAAAACAGTCACTGCATGGATCCATTGAATGTTGCTGCTTATTCTCATTAACT
CACAGATATTCAAGAACATCGTACTCTCAACCAGGCTAAAGCAAAGAGGGTACATTAGCCG
ACAAGTTCACTAGCTGAGTGGAACACGTATATTAATGGAGATGACTCTGGTCATGATGATTAGG
ACAATTATCATGACGTTATCATTGATCATGACCAGTGTAGTAAATAGATAGCTAACAAATAATGT
ATTACTAATTATGAAGCAATGGTCATTGCA

DOMAIN 1F-2 (2nd part of domain f)

Fig. 8d

GTGGCATGCTGTATCCATGGAATGCCTACATTCCCCACTGGCACAGACTCTACACCCTCCAATT
 GAGCAAGCTCTAAGAACATGGCTCTAGTGTAGCAGTACCTACTGGGACTGGACAAAGCCAATA
 CATAATATTCCACATCTGTCACAGACAAAGAATACTACGATGTCTGGAGAAATAAGTAATGCCA
 AATCCATTGCCCGAGGGTATGTCCCCTCACACGATACATACACGGTAAGAGACGTCCAAGAAGGC
 CTGTTCCACCTGACATCAACGGGTGAACACTCAGCGCTCTGAATCAAGCTTTGGCGCTGGAA
 CAGCACGACTACTGCGATTGAGTCCAGTTGAAGTCATGCACAACACAATCCATTACCTAGTG
 GGAGGACCTCAAGTCTATTGTCATCCCTCATTATGCTCATATGATCCGATCTCTTCATA
 CACCACTCCTTGTAGACAAGGTTGGGCTGTCTGGCAGGCTCTCAAGAAAAGAGAGGGCCTTCCA
 TCAGACCCTGACTGCCTGTTAGTCTGACTCAGAACATGAGGCCATTACGAAATT
 AACCATAACCAAGAACATGCAGTCCAAATGATGTTCAAGTACGAACCTCTGGGT
 TACAGATACGACAATCTGAAATCGGTGGCATGAATTGTCATGAAATTGAAAAGGAAATCAAAGAC
 AAACAGCACCATGTGAGAGTGTGAGGGTTCCCTCAGGAATTAGAACCTCAGCTGATGTC
 CAATTCCAGATTGTAACATCAGAACAGATTGTCACCAGGAGGCCAAATCTCGTCTGGGGGG
 ACTAAAGAGATGCCCTGGGCTATAACGTTATTCAAGTACGATATTACCCATGCTCTCATGAC
 GCACACATCACTCCAGAACAGTATTCCATCCCTCTGAACCATTCTCATCAAGGTGTCAGTGACA
 GCCGTCAACGGAACAGTTCTCCGGCTCAATCCTGCATGCACCAACCATTATCTATGAACCTGGT
 CTCGGTG

INTRON 1F-2/1G-1 (SEQ ID NO:144)

GTCTCGGTGAGTTATTAAAAGAAACAAATATTACCATTACCATTGTTAATCACAAAATGAGTG
 AGATATCTTATATCACTGGTACACTACTGATATTGCAATGAAATTACTATTTCCAGGTAC
 GCTTCAACCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCATCATGCTTTCTGT
 AAAACATAAAACACCAATTAAACAATGTTCTTAGTGTGTTGACTCCCTTCACTGCAACGCCT
 ACATAATCAAAGTGTCTGTTTCCAAACTTTCCAGTTAGTGTGAAGACTAAAAGTTAAATA
 AGCATTCACATAACTTCTAAGAGCAACTGGGACCATGCAGTTACGTATTGATATTCTGTGAGAGT
 GAAGCAAAACACTGTTTCAAGCTTAGGTTATCAATCAAATGTCCAATAGTTCATGTTATCGA
 AAAGGCAGCGAAGGATAAGAGGGCTCCGAGACATCTGTCTATTCTCGTGTTCATATGATAC
 GAGGAGCTTCCATTACATTGACCTTATCATTAAAGACATACATGGAACATTTCATTTACA
 GTTAAAGTGAACCACTTCAGGTTCAACTTCAGGTTCAACTTCAGGTTCTGTTGTTATGAG
 CCGACTGAAATAGAGTGCCTTACTTCAGGTTCTGTTGTTCTGTCATGTTGTTCTT
 CAGTGTGCATAGTACACGCCAGTATAGAACACACGAACATTGTCCTTACTTAATAGATTCTGAAAC
 TATTATGTGGAAAGTTGGCAGGCTATAGTAACATCCTGGAAAATTATCATGTATCCTCTGTTG
 TCATAATTAG

DOMAIN 1G-1 (1st part of domain g)

ACCATCACGAAGATCATCATTCTTCTATGGCTGGACATGGTGTAGAAAGGAAATCAACACAC
 TTACCACTGCAGAGGTGGACAATCTCAAAGATGCCATGAGAGGCCGTATGGCAGACCAAGGTCCAA
 ATGGATACCAAGGCTATAGCAGCGTTCCATGGAAACCCACCAATGTGCCCTATGCCAGATGGAAAGA
 ATTACTCGTGTGACACATG

INTRON G1-1/1G-2 (SEQ ID NO:145)

GTATGTATTCCTTCACTGGTGGTCGCTGACTGCCAACACATACTTGTAAATTATTCATGAAAGTATA
 ATAGTTGTTGAAAGTATATTATAACCATCTGCACAAGCGTCACGAATTTCACCACAAAGCT
 TCAAAACGCCAAAACATTCTAAATAGCGATATATTGTTAAAAGACCAAAATATAGCCTTACAACA
 ATAGATTATTTAATAAGACCAGTCAGTGCATGCAAATCGATTGAAACTTGTGAAATAAAATATT
 TATGTACTAACTGCCAATCTCATAATACTGCCTGGATGTGCTTCTTACATTGCGTGGAG
 CTTCAACTCCAATGCATAAGCTTAAAATAATCATAAACACAAATAGCCACAGAGGGCACGA
 TCCCTCCAGGCCAGGCTTATTGTCATAGAATATCGCTATTAGAATGTTTGACGTTT

Fig. 8e

TGAAGCTTGTGGTAAAAATTCTGTATGTTATGCGTGGTATTATGTAAGATGAAAATAAATAT
 ATCTTTCAAACAAGATTAGTATTTGAAGACTTCTATGAATAAAATTACACTTATGTGTTAGGT
 TATTGGTCAGTGCCTGTGGTATTCCCTTCAATTGTTGTTCTTGTCAATTCTGA
 ATAGTTATCCTACTGTGGATAGTCTATATGAGAACGTTGAAAGAATAATACAATTCTAATGGATT
 GCAACTTCTTAACCTTATTGCAACTGCCACGTTGGTACGTTCTATGCCGTATCAAGC
 ATACGAGTGTACATGTATGCCAAAACGCTGCAAATAAAATTAAAGAAGTTGCAATCCATAAGAAT
 TTCAATGTTCTTCATCATCACATCAACTCTAAAGAATGCCTATAAAACAATCAACAAACGTACAA
 TAGTACATTACCGGATCTCGCAGCATGACCACGTCGATATCTAAACAATATCACTATCCATTAAATA
 GGATCAAGAGTAGGTACAGACATGTTAGTTATAAAACTCTTAAAAAGTAGGGAACTGGAA
 TTTCAAGGTCAATAACAAACTAATGATAATAACAATTGGTCCAAATAATAACAATTGGTCCAA
 CTAATTGTATCTTACAAAGAAGAAATTGAGTGAACAATTACCCGGTATTATTACCTAAACCG
 TTTCTCTGCTGTTATGGTGCCTGAAAGAAGAAATGGTAAGAAACGGAAATTGACATTTGCGT
 CAGTGGTGCCTAATGCCCAATTGTTGCCAAACACTGATTGATTGCTGAGGCATCGTCATACG
 CGTCTACCTATGGAATTGATGCAGTCTGTCCTTCCACCAACGCGCTGGACAAGTTCATCT
 AGCGTGGCTGGTGGCCTTACGTTGACGCACACGTCGGCCAAGATGTCAGACATTTCAATG
 GCCAGGGCTCATTGCTGGTCAGGGCATCCTATGGATTGTCAGGAGGTTACAAAGTACGTTCTTCATGTT
 GGTGAAGAGAATATCAAGGTCTTAAGGGATTGTTCTATAATATTGATTAAAGAAGTTGA
 TATTATCTGCATCCTTCCAAGAAATTGCAAATGTCACACACTATTGCTTGATAATGTTTG
 GGGAAATAAAACTGTCCAGGACTGCTAAATAGTAATTATTGCTACTTTAG

DOMAIN 1G-2 (2nd part of domain g)

GCATGGCTACTTCCCCCACTGGCACAGACTGTACACAAAACAGATGGAAGATGCCTGACCGGCC
 ATGGTGCCAGAGTCGGCCTTCTACTGGGACGGACACTGCCTTACAGCTTGCAACTTTG
 TCACAGATGAAGAGGACAATCCCTTCCATCAT

INTRON 1G-2/1G-3 (SEQ ID NO:146)

GTGAGTTCACGTAAGCCTACGAGATCAACATTACTCCTAACAGCCACGGCATCATGTACCGATAT
 ATCACAAACAAAAGTATTCAAAGCTTAAACACGATATGTATGGTTCAAGAACATCATTAAAC
 AAGGACATGAGTCTGAAATAAACATGACTTGACACCGTTGTCAGTTGTTCTCATGGT
 GAACCTGTGAAACAAACCTTCAAACCAAAAGATGCCTATTAAATATTGTTAATTCCATGAATTAGG
 AGATACACACATCTACTGTCATT.....AATAACCGCTTC
 CAGCATGAAAACACAATATGATTATCTCAATTCTACCAATTACTAATTATAATTGACTGGCATT
 TTTGACGACGCGTAAACATCGCTGCTTACAGACTGCACTGCGGTAACGTGACGTTTCATGAC
 GTCACTACATTCTATTCAAACATTCCACAGAACAGAGCGAGACCACGGCCGTGATGGGTTCTGGGC
 AGATGATTACCCAAGTATATATTATAATAACTTGACTGCTGCCTGAATAATGTTGACACATGAC
 AACGAATTGTAAGCGTAAGAACCGTGAATACTGTGAATAGTGTGAGGGGTGTTGCTGAGAGT
 TAACCACCGTTAATTGCAAAATTCCGAATACTTGCAATTGCAAGAACAGAACATTGCAATTCTTA
 CTCCTGTGAATGGACTCATGTTATTAGCAGCGGTTATTGAGGTTTGATCACCTCTAAATAGAC
 AATCAGGATGCGCAAACCGAAAATTATAGCAGAACATCTGTAATTCAAGATGGCCTGCCTGTGAA
 AATATGCTGCGAGGTTCAAGAACACTTTCCCTTCGATCATGGCCTGTTGCTCTGAATCTGGTC
 TTTCAGAGGATCCCTGCTTTAAAACAAAGTCCTCCAACTCACTTATATTATGTTTTAA
 TTATTATAGTTAATATGAACAACAAATCATATTACACATTATTTTCAG

DOMAIN 1G-3 (3rd part of domain g)

GGTCACATAGACTATTGGGAGTGGATACAACACTCGGTGCCCGAGACAGAACAGTTGTTCAATGATCCA
 GAGCGAGGATCAGAACATGTTCTACAGGCAGGTTCTGGCTTGGAGCAGACAGAT

Fig. 9aPrimary structure of the KLH1 proteinDOMAIN B

GLPYWDWTEPMTHI PGLAGNKTYVDSHGASHTNPFHSSVIAFEENAPHTKRQIDQRLFKPATFGHH
 TDLFNQILYAFEQEDYCDFEVQFEITHNTIHAWTGSEHFSMSSLHYTAFDPLFYFHHNSVDRLWA
 VVQALQMRRHKPYRAHCAISLEHMHLKPFSSPLNNNEKTCHANAMPNKIYDYENVLHYTYEDLTF
 GGISLENIEKMIHENQQEDRIYAGFLLAGIRTSANVDIFIKTTDSVQHKAGTFAVLGGSKEMKWGF
 DRVFKFDITHVLKDLDLTADGDFEVTVDITEVDGTKLASSLIHASVIREHARGKLN

DOMAIN C

VKFDKVPRSRLIRKNVDRSLPEEMNELRKALALLKEDKSAGGFQQLGAFHGEPKWCPSPASKKFA
 CCVHGMCSVFPWHRLLTVQSENALRRHGYDGALPYWDWTSPLNHLPELADHEKYVDPEDGVEKHNP
 WFDGHIDTVDKTTTRSVQNKLFEQPEFGHYTSIAKQVLLALEQDNFCDFEIQYEIAHNYIHALVGG
 AQPYGMASLRYTAFDPLFYLHHSNTDRIWAIWQALQKYRGKPYNVANCAVTSMREPLQPGFLSANI
 NTDHVTKEHSVPPNFVFDYKTNFNEYDTLEFNGLSISQLNKKLEAIKSQDRFFAGFLLSGFKKSSL
 VKFNICTDSSNCHPAGEFYLLGDENEMPWAYDRVFKYDITEKLHDLKLHAEDHFYIDYEVFDLKPA
 SLGKDLFKQPSVIHEPRI

DOMAIN D

GHHEGEVYQAEVTSANRIRKNIENSLGELESLRAAFLEIENDGTYESIAKFHGSPGLCQLNGNPI
 SCCVHGMPTFPWHRLYVVVVENALLKKGSSVAVPYWDWTKRIEHLPHLISDATYYNSRQHHYETN
 PFHHGKITHENEITTRDPKDSDLFHSDYFYEQVLYALEQDNFCDFEIQLEIILHNALHSLLGGKGKYS
 MSNLDYAAFDPVFFLHHATTDRIWAIWQDLQRFRKRPYREANCAIQLMHTPLQPFDKSDNNDEATK
 THATPHDGFEYQNSFGYAYDNLELNHYSIPQLDHMLQERKRHDRVFAGFLLHNIGTSADGHVFVCL
 PTGEHTKDCSHEAGMFSILGGQTEMSEVFDRLYKLDITKALKKNGVHLQGDFDLEIEITAVNGSHL
 DSHVIHSPTILFEAG

DOMAIN E

TDSAHTDDGHTEPVMIRKDITQLDKRQQSLVKALESMKADHSSDGFQAIASFHALPPLCPSPAAS
 KRFACCVHGMATFPQWHRLYTVQFQDSSLRKHGAVVGLPYWDWTLPRSELPELLTVSTIHDPETGRD
 IPNPFIGSKIEFEGENVHTKRDINRDRLFQGSTKTHHNWFIEQALLALEQTNYCDFEVQFEIMHNG
 VHTWVGGKEPYGIGHLYASYDPLFYIHHQSQTDRIWAIWQSLQRFRGLSGSEANCNAVNLMKTPLKP
 FSFGAPYNLNDHTHDFSKPEDTFDYQKFGYIYDTLEFAGWSIRGIDHIVRNQEHSRVFAGFLLEG
 FGTSATVDFQVCRTAGDCEADGYFTVLGGEKEMPWAFDRLYKYDITETLDKMNLRHDEIFQIEVTI
 TSYDGTVLDGLIPTPSIYDPAH

DOMAIN F

HDISSHLSLNKVRHDLSTLSERDIGSLKYALSSLQADTSADGFAAIASFHGLPAKCNDSHNNEVA
 CCIHGMPTFPWHRLYTLQFEQALRRHGSSVAVPYWDWTKPIHNIPLHFTDKEYDVWRNKVMPNP
 FARGYVPShDTYTVRDVQEGLFHLTSTGEHSALLNQALLALEQHDYCDFAVQFEVMHNTIHYLVGG
 PQVYSLSSLHYASYDPIFFIHHSFVDKVWAWQALQEKRGLPSDRADCAVSLMTQNMRFHYEINH
 NQFTKKHAVPNDVFKYELLGYRYDNLEIGGMNLHEIEKEIKDKQHHRVFAGFLLHGIRTSADVQF
 QICKTSEDCHGGQIFVLGGTKEMAWAYNRLFKYDITHALHDAHITPEDVFHPSEPFFIKVSVTAV
 NGTVPASILHAPTIYEPGLG

Fig. 9b

DOMAIN G

DHHEDHSSSMAGHGVRKEINTLTTAEVDNLKDAMRAVMADHGPNQYQAIAAFHGNPPMCPMPDGK
NYSCCTHGMATFPWHRLYTKQMEDALTAH GARVGLPYWDGTTAFTALPTFVTDEEDNPFHGHID
YLGVDTTRSPRDKLFDPERGSESFFYRQVLLALEQTD

Fig. 10aGenomic sequence of the KLH2 gene

DOMAIN 2B

```

GGCCTGCCCTACTGGGATTGGACCATGCCAATGAGTCATTCGCCAGAACTGGCTACAAGTGAGACC
TACCTCGATCCAGTTACTGGGGAAACTAAAAACCAACCCCTTCCATCACGCCAAGTGGCGTTGAA
AATGGGTGTAACAAGCAGGAATCCTGATGCCAACTTTTATGAAACCAACTACGGAGACCACACT
TACCTCTTCGACAGCATGATCTACGCATTTGAGCAGGAAGACTTCTGCGACTTGAAGTCCAATAT
GAGCTCACGCATAATGCAATACATGCATGGTTGGAGGCAGTAAAAGTATTCAATGTCCTCTCTT
CACTACACTGCTTTGATCCTATATTTACCTCCATCACTCAAATGTTGATCGTCTCTGGGCCATT
TGGCAAGCTCTCAAATCAGGAGAGGCAAGTCTTACAAGGCCACTGCGCCTCGTCTCAAGAAAGA
GAACCATTAAAGCCTTTGCATTCACTGGGGAACTGAACAACAACGAGAAAACGTACCACAACCT
GTCCCCACTAACGTTATGACTATGTGGGAGTTGCCTATCGATATGATGACCTTCAGTTGGC
GGTATGACCATGTCAGAACATTGAGGAATATATTACAAGCAGACACAACATGATAGAACCTTGCA
GGATTCTCCTTCATATATTGGAACATCAGCAAGCGTAGATATCTTCATCAATCGAGAAGGTCT
GATAAATACAAAGTGGGAAGTTGTAGTACTTGGGATCAAAGAAATGAAATGGGGCTTGAT
AGAATGTACAAGTATGAGATCACTGAGGCTCTGAAGACGCTGAATGTTGAGTGGATGATGGGTT
AGCATTACTGTTGAGATACCGATGTTGATGGATCTCCCCATCTGCAGATCTCATTCCACCTCCT
GCTATAATCTTGACGTGGTCAGAG

```

INTRON 2B/2C (SEQ ID NO:147)

```

GTATTTAAAAAAGTAATAAAACCATAATTCGAATGCGTTTATGAAATATCGTGTGACTGGTTCT
TTAGTTACATGGAGTGTAAACAACATGCTCCATCAGTTGACATATACTGCTCACACAAAGTAAGGG
ATATTTGATAATGATAACAAATATAATCAAAGCGTTATACTATCAAGACTTATTACATAATTAC
AGGTGAAGGGAGGTGTGATCGTGTTCAGTGTGAGGCAACTACCAAAATCACATGACTTCAATAACGGGTTGG
TGCAGAAAGATGATGTTAGGCATGGGTGCAATCACCACATGACTTCAATAACGGGTTGG
ACCACCTCGAGCGACGATGCAAGCAGTAGAGCGTCTACGCATGCTCTGATAAGGCACCAATCTG
TTCCTGGGAATCAGTCGCCACTCCTCTTGAGTGCCACGCTCATTCTGCTACGGTCTGGGTAC
CTGCTATCGGGTCTTGATCCGTATCCAAGGATGTCACACACATGTCAGTCAAGGTGAGAGGGTCGGGGA
ACATCGCTGCCACGTAAGGTCTGAATTGATGCCGTTGAAAGTGAAGCTCTGACAAACCTGAGCAT
GGTAGCTCTGACGTTGTCGTCCTGAAAGATGAATCCAGCTCCATGACAGCGAGCAAAGGGCAGGA
CGTGGACTCGTCACTGAACATGACATTAGCCAACCTGGCAGTGTCCACCGCTGATGTTGGCGAGA
CCATTCCAGTCGAGCTCTCGGTGTCTGGCTTCATCGATAACACGACGTAAGGTCTGCGGGCGTG
CAAGACGGCTCATGCAAGCGATTTCGGATTGTCTGGGTGCTAACTCTGATCCCAGGTGCCTGCTG
AAGTTGATGCTGGATCTGTGTGGCATTGAGATGGCGATTCCCTAGGACTGTGGAGATGATGAATCG
ATCTTGACTTATGGTGGTGCACATTAGGACGTCGGGTGCTGCTATCCTGCACCTTCCAGTTGT
TCGGTGACGCTCTGGTACCCGGCTGATTACTGACTGAGAATATCCATCTGCCGTGCGACATGAGCC
TGTGTTGGCCAGCCTGAAGCATTGCAATGCCAGAGACGCTCTCAAAAGTCATTGACGCATGG
TTTCTGTTACAAATGACAGCGTAAACAGTTGGTGTCTTATGCTTCAAGAGCATGAAA
AACACGTTCTATGGTCGTGCACACCTACATGACAAGTGTGAAAGTGAATTGCAACCCCTTGTG
TGTCGGATGCACACTCTGTTACGTACTGCGATTGGCGTCTAAACATGTTGGCGTCTAA
ACATGTTCTGCATGATTCAATATACTATTTGTCATATTCTGGCATCAAACCAAAACTACAGTG
AAATATATTCATATCCCTACTTTGTCAGTAGTATAGATCACTGCAAGACAAACATATAGACAA
TGCAGTTACACCGTCAACAATCCCAGTCATTAATTGATGACACTTCCACACATAGTGTAGTGA
TTGTAATTCAACTGTACACACTTTCCGTGAACATTGAGGATCTATGACTAAATATATAACAT
TAGTATACGTGCAGTTGTATCGCTACGACATTGTTGTAACCTTTGTTAACATTAAACAG

```

Fig. 10b

DOMAIN 2C

CTGATGCCAAAGACTTGGCCATAGCAGAAAATCAGGAAAGCCGTTGATTCTCTGACAGTCGAAG
 AACAAACTCGTTGAGGCGAGCTATGGCAGATCTACAGGACGACAAAACATCAGGGGGTTCCAGC
 AGATTGCAGCATTCCACGGAGAACCAAAATGGTGTCCAAGCCCCGAAGCGGAGAAAAAATTGCAT
 GCTGTGTTCATGGAATGGCTGTTCCCTCACTGGCACAGATTGCTGACAGTCAAGGAGAAAATG
 CTCTGAGGAAACATGGATTACTGGTGATTGCCATTGGGACTGGACTCGGCCATGAGCGCCC
 TTCCACATTTGTTGCTGATCCTACTTACAATGATTCTGTTCCAGCCTCGAAGAAGATAACCCAT
 GGTATCATGGTCACATAGATTCTGTTGGCATGATACTACAAGAGCTGCGTGATGATCTTATC
 AATCTCCTGGTTCGGTCACTACACAGATATTGAAAACAAGTCCTCTGGCCTTGAGCAGGACG
 ATTTCTGTGATTTGAGGTACAATTGAAATTGCCATAATTACATGCTCTGGTTGGTGGTA
 ACGAACCATACAGTATGTCATCTTGAGGTATACTACATACGATCCAATCTCTTCTGCACCGCT
 CCAATACAGACCGACTTGGGCCATTGGAAGCTTGCAAAAATACCAGGGAAACCATAACA
 CTGAAACTGTGCCATTGCATCCATGAGAAAACCACCTCAGCCATTGGTCTGATAGTGTCAA
 ATCCAGATGACGAAACTCGTGAACATTGGCTCCTTCCGAGTCTCGACTACAAGAACAACTTCG
 ACTATGAGTATGAGAGCCTGGCATTAAATGGTCTGCTATTGCCAACACTGGACCGAGAGTTGCAGA
 GAAGAAAGTCACATGACAGAGTCTTGAGGATTCCCTCATGAAATTGGACAGTCTGCACTCG
 TGAAATTCTACGTTGCAAACACAATGTATCTGACTGTGACCATTATGCTGGAGAATTCTACATT
 TGGGAGATGAAGCTGAGATGCCTGGAGGTATGACCGTGTACAAGTACGAGATAACACAGCAGC
 TGCACGATTAGATCTACATGTTGGAGATAATTCTCTTAAATATGAAGCCTTGATCTGAATG
 GCGGAAGTCTGGTGGAGTATCTTCTCAGCCTCGGTATTTCGAGCCAGCTGCAG

INTRON 2C/2D (SEQ ID NO: 148)

GTATGTTTAAATGTCACTTATCCGTGATCTGAAATGAAGTTAGCAATTCACTTATCAACTGTT
 GGCTGTACTGTTCACTGCGAGTTTACTTAGGTTGGATTAATTAAAATTCAAGCTCATAATG
 TTTTGATTCAACTTTGTTATTATTCAAACAG

DOMAIN 2D

GTTCACACCAGGCTGATGAATATCGTGAGGCAGTAACAAGCGCTAGCCACATAAGAAAAATATCC
 GGGACCTCTCAGAGGGAGAAATTGAGAGCATCAGATCTGCTTCCCAAATTCAAAAAGAGGGTA
 TATATGAAAACATTGCAAAGTCCATGGAAAACCAGGGACTTGTGAACATGATGGACATCCTGTTG
 CTTGTTGTGTCATGGCATGCCACCTTCCCCACTGGCACAGACTGTACGTTCTCAGGTGGAGA
 ATGCGCTCTTAGAACGAGGGCTGCAGTTGCTGTTCTTACTGGGACTGGACCGAGAAAGCTGACT
 CTCTGCCATCATTAAATCAATGATGCAACTTATTCAATTACGATCCAGACCTTGATCCTAA
 CTTTCTCAGGGACATATTGCCTCGAGAACATGCTGTGACGCCAGAGATCCTCAGCCAGAACTAT
 GGGACAATAAGGACTTCTACGAGAACATGTCTGCTGGCTTTGAGCAAGACAACCTCTGACTTTG
 AGATTCAAGCTTGAGCTGATACACAAACGCCCTCATTCTAGACTGGAGGAAGGGCTAAATACTCCC
 TTTCGTCTCTGATTATAACCGCATTGATCCTGTATTTCCTCACCATGCAAACGTTGACAGAA
 TCTGGCCATCTGGCAGGACTTGCAAGAGATAGAAAGAACCATACAATGAGGCTGACTGCGCAG
 TCAACGAGATGCGTAAACCTCTCAACCATTAAATAACCCAGAACCTAACAGTGATTCCATGACGC
 TTAAACACAAACCTCCCACAAGACAGTTGATTATCAAACCGCTTCAGGTACCAATATGATAACC
 TTCAATTAAACCACTTCAGCATAACAAAGCTAGACCAAACATTAGGCTAGAAAACAACACGACA
 GAGTTTTGCTGGCTTATTCTCACAAACATTGGACATCTGCTGTTGAGATATTATATTGCG
 TTGAACAAGGAGGAGAACAAACTGCAAGACAAAGGGCTCCTCACGATTCTGGGGGAGAAA
 CAGAAATGCCATTCCACTTGACCGCTGTACAAATTGACATAACGTCTGCTCTGCATAAAACTTG
 GTGTTCCCTGGACGGACATGGATTGACATCAAAGTTGACGTCAAGAGCTGTCATGGATCGCATC
 TTGATCAACACATCCTCAACGAACCGAGTCTGCTTTGTTCCCTGGTGAACGTAAGAATATATT
 ATG

Fig. 10c

INTRON 2D/2E (SEQ ID NO:149)

GTTATAAAGCAGTATATTCTCTAAAAAGTAGGGAACTTGAATTCAAGGTAAATAACATAA
 CTACCTCAACGGCACAAATATCCATATGATGCCCTGGCCAGCAATGAGGCCTGATCTTCCCCT
 TAAAAATGTCTGGAACATCTGGCAAACGTGTGCGTCACGTAAAAGCCACCAGTCACGCTAGA
 TGAACTTGTCCAGGCCTGGGAGAATGGACAGACTGCATCAATTACCATAGTAGACTCATT
 TGCAGCGAATCAGTCAGTGTGACCAATAACGGGGCATTACGCACTACTGACGAAAACAATGT
 CAATTCCGTTCTTACCCATTCTTACGGACCATAACAGCAAGAGAACTGNNTAGGTAA
 TGAAATACCGGTGAATTATTGTTACTGGATTCTTGTAAAGATAACAATTAGTTGGGACCA
 ATTATTATTATCATTAGTTGTTATTGACCTGAAATTGAAAGTTCCTCTACATTTTAAGGAGT
 TTATTTGATTGACAATGAAATGTAAGAAAAGAGCAAATCGTAAAATACGTTAAAATTATTCTTA
 AACATCAGTCTACTTCAGTTAAATTGCCAGTAACACGTGTTATGATGTTCCGTTCTCT
 TTGTTTTAGCATTCAACTTATTGATATAACGTTACTGTTAGATTACATCAAACATGCAG

DOMAIN 2E

ATGGGCTTCACAACATAATCTTGTGCAAAAGAAGTAAGCTCTTACAACACTGGAGAACATT
 TTTTGAGGAAAGCTCTCAAGAACATGCAAGCAGATGATTCTCCAGACGGATATCAAGCTATTGCTT
 CTTTCCACGCTTGCCTCTTGTCCAAGTCCATCTGCTGCACATAGACACGCTTGTGCCTCC
 ATGGTATGGCTACCTTCCCTCAGTGGCACAGACTCTACACAGTCAGTTGAAGATTCTTGAAAC
 GACATGGTTCTATTGTCGGACTTCCATATTGGATTGGCTGAAACCGCAGTCGACTCCCTGATT
 TGGTGACACAGGAGACATACGAGCACCTGTTTACACAAAACCTTCCAAATCCGTTCTCAAGG
 CAAATATAGAATTGAGGGAGAGGGAGTAACAACAGAGAGGGATGTTGATGCTGAACACCTCTTG
 CAAAAGGAAATCTGGTTACAACAACAGGTTGCAATCAGGCACTATATGCACTAGAACAGAAA
 ATTACTGTGACTTTGAAATACAGTTGAAATTTCGATAATGGAATTCTATTGTTCTATATCCACCA
 CAAAGACCCATTCAATAGGTCTTCATTACGCACTACGATCCACTGTTCTATATCCACCA
 CGCAGACAGATCGCATTGGCTATCTGGCAAGCTCTCCAGGAGCACAGAGGTCTTCAGGGAAAG
 AAGCACACTGCGCCCTGGAGCAAATGAAAGACCCCTCTCAAACCTTCAAGCTTGGAAAGTCCCTATA
 ATTTGAACAAACGCACACTCAAGAGTTCTCCAAGCCTGAAGACACATTGATTATCACCAGTCGGGT
 ATGAGTATGATTCCCTCGAATTGTTGGCATGTCGTTCAAGTTACATAACTATATAAACAC
 AACAGGAAGCTGATAGAGTCTCGCAGGATTCTTAAAGGATTGGACAATCAGCATCCGTAT
 CGTTGATATCTGCAGACAGACCAGAGCTGCAAGAAGCTGGATACTCTCAGTTCTCGGTGGAA
 GTTCAGAAATGCCGTGGCAGTTGACAGGCTTACAAGTACGACATTACAAAACGTTGAAAGACA
 TGAAACTGCGATACGATGACACATTACCATCAAGGTTCACATAAAGGATATAGCTGGAGCTGAGT
 TGGACAGCGATCTGATTCCAACTCCTCTGTTCTGAAAGAAGGAAAGC

INTRON 2E/2F (SEQ ID NO:150)

GTATGTATCTCATGTTCTCAAATAATTGATTTCATGCCCTACTATAAAGCACAGTTATTGT
 TCAGTGCCAGTAACCGTTATTACGTAATGTTACAGGCTATTATAATCAAAAATACATTACCGA
 TATTGTTACCAACACAATTATATCATTGTCAAATCTACCCCCATTACCTGCGTTTGAATTGTA
 ACCTTCTGACAAAAATGAATTAGCAAGAGCTCTGATGAAGAACATAATGAACAAACACCTATCTTC
 TTCTTCAATGACGGTTAACATACAATGCACAATGTAACAAAATATATATATATATAATT
 ATATCTACAGTTAATGCAAATGACTCCACTAATTCAAGGAAACACATTTCAG

DOMAIN 2F-1 (1st part of domain f)

ATGGGATCAATGTACGTACGTTGGCGTAATCGGATTGATGAACTATCTGAACTCACCGAGA
 GAGATCTGCCAGCCTGAAATCTGCAATGAGGTCTACAAAGCTGACGATGGGTGAACGGTTATC
 AAGCCATTGCATCATTCCACGGCTCCGGCTTGTGATGATGAGGGACATGAG

Fig. 10d

INTRON 2F (SEQ ID NO:151)

GTAAAATAAAACGTCCAGTCATCGGAAACCCGCCAGATATGGGTTTTCTATTAAACAAA
 AAAGCAGAGACAAAAAGATTATAAAGTCACATTAACTTGATATCAGATCAAATAGTTGGCTAG
 TTAGTGCTCTATATCCCTCAAATCCTCGAATCTTAAGCCTCGTGAATTTGACAACAGAGAA
 GACTTAGTAGCCCAGACTTCCCTATTTTCTGAAAATCTTAATACGGATATTAAATGGATTC
 ATTCTGCAACCTACAACCAGCCCATACTGTTATTATTCAG

DOMAIN 2F-2 (2nd part of domain f)

ATTGCCTGTTGATCCACCGAATGCCAGTATTCCCACACTGGCACAGGCTTACACCCGTCAAATG
 GACATGGCTCTGTTATCTCACGGATCTGCTGTTGCTATTCCATACTGGACTGGACCAAACCTATC
 AGCAAACGTGCTGATCTCTCACCGCCCTGAATATTACGATCCTTGGAGGGATGCAGTTGTCAAT
 AATCCATTGCTAAAGGCTACATTAATCCGAGGACGCTACAGGTTAGGGATCCTCAGGACATT
 TTGTACCACTTGCAGGACGAAACGGGAACATCTGTTGTTAGATCAAACCTTTAGCCTTAGAG
 CAGACAGATTCTGTGATTTGAGGTTCAATTGAGGTCGTCATAATGCTATTCACTACTGGTG
 GGTGGTCGACAAGTTATGCTTTCTCTCAACACTATGCTTCATATGACCCAGCCTCTTATT
 CATCACTCCTTGTGACAAAATATGGCAGTCTGCAAGCTGCAAAGAAGAGAAAGCGTCCC
 TATCATAAAAGCGGATTGTGCTCTAACATGATGACCAAACCAATGCCACATTGCACACGATTTC
 AATCACAATGGATTCACAAAATGCACGCAGTCCCCAACACTCTATTGACTTTAGGACCTTTTC
 TACACGTATGACAACCTAGAAATTGCTGGCATGAATGTTAATCAGTTGAAGCGGAAATCAACCGG
 CGAAAAAGCCAACAAAGAGTCTTGCCGGGTTCTACATGGCATTGGAAGATCAGCTGATGTA
 CGATTTGGATTGCAAGACAGCTGACGACTGCCACGCATCTGGCATGATCTTATCTTAGGAGGT
 TCTAAAGAGATGCACTGGGCCTATGACAGGAACCTAAATACGACATCACCAAGCTTGAAGGCT
 CAGTCCATACACCCCTGAAGATGTGTTGACACTGATGCTCCTTCTTCATTAAAGTGGAGGTCCAT
 GGTGAAACAAGACTGCTCTCCATCTCAGCTATCCCAGCACCTACTATAATCTACTCAGCTGGT
 GAAG

INTRON 2F-2/2G (SEQ ID NO:152)

GTGAGAGAAACTATAATAGTGTATGCGGAAAAATGTGCTCATATCATGACTCTGTTGGCCGGT
 GGTTGCTCTCCTCTCCTCCACCACCAACCGGTACCTCCACCTGTCAGGGCATCAATGTACCATG
 AAAATGTCTACAATACTAGGCCTCCTGTAGAACGACGTAAGATTACATGGCCGGTTGTAACTAG
 TTTAAAGTGTTCACAGTAACCAAAACAGTCTCTAAAGATTAATGTCGTTAAAATTAAATGCC
 ACATTTCAACTGACATATTCTTGCATTAAGTACAAATGAAGTAGTATAAATTATCCACAAATAG
 CGTGATGCACCACAAATAAAACCGAGTGCTTTGGCATCCCCACTTGTCTGGCATGATCAC
 ATCATAGATCTGTTCATGAAGATACTGTTGGATGCTTTCCAATATGCCCAATCTGTTAAAT
 TATTACACGACCGCAGTGTGACTTCATCACTCAGATCTTACAATGTGTTGTAACGTTACA
 ATTAGCGTTATGATTGAAATATTACCCCTGCTACGTTAAATCACATTCACTCACTCATGATGT
 ACTTTACAGGTACCGATGATCACGGCTCAG

DOMAIN 2G-1 (1st part of domain g)

ATCATATTGCTGGCAGTGGAGTCAGGAAAGACGTGACGTCTTACCGCATCTGAGATAGAGAAC
 TGAGGCATGCTCTGCAAAGCGTGTGGATGATGGACCCAATGGATTCCAGGCAATTGCTGCTT
 ATCACGGAAGTCCTCCATGTGTACATGCCTGATGGTAGAGACGTTGATGTTACTCATG

INTRON 2G-1/2G-2 (SEQ ID NO:153)

GTCAGTATTCTCCAATATGTTGACTAGTGTCTGCTCATGTATCAACTATTTAGGCAACGTTT
 TGATTGTTATGGTATTTCATGATATGATTGCTACCTCTACCCAAACAAAATGTTTA

Fig. 10e

TCAACAATTGTTGAGTTTAATGCAAGAAAATTATCAGGAGTAGCGTGCAAAATGACTGGAAGG
CATGGTGTACTTCTGTGTACATACAAGTGGTAATGCCTTATTGAACCTCGTAATCACTCGTTTC
AG

DOMAIN 2G-2 (2nd part of domain g)

GAATGGCATCTTCCCTCACTGGCACAGACTGTTGTGAAACAGATGGAGGATGCACTGGCTGCGC
ATGGAGCTCACATTGGCATACCATACTGGATTGGACAAGTGCCTTAGTCATCTGCCTGCCCTAG
TGACTGACCACGAGCACAAATCCCTCCACCAC

INTRON 2G-2/2G-3 (SEQ ID NO:154)

GTCAGTATTCTCCAATATGTTGACTAGTGTCTTGCTCATGTATCAACTATTTAGGCAACGTTT
TGATTGTTATGGTATTTCATGATATGATTTATTGCTACCTCTATACCAAACAAAATGTTTA
TCAACAATTGTTGAGTTTAATGCAAGAAAATTATCAGGAGTAGCGTGCAAAATGACTGGAAGG
CATGGTGTACTTCTGTGTACATACAAGTGGTAATGCCTTATTGAACCTCGTAATCACTCGTTTC
AG

DOMAIN 2G-3 (3rd part of domain g)

GGACATATTGCTCATCGGAATGTGGATACATCTCGATCTCCGAGAGACATGCTGTTCAATGACCCC
GAACACGGGTCAAATCATTCTTCTATAGACAGGTTCTTGGCTCTAGAACAGACAGACTTCTGC
CAATTGAAAGTTCACTGGAAATAACACACAATGCAATCCACTCTTGGACTGGAGGACATACTCCA
TATGGAATGTCATCACTGGAATATACAGCATATGATCCACTCTTATCTCCACCATTCAAACACT
GATCGTATCTGGCCATCTGGCAGGCCTCCAGAAATACAGAGGTTTCAATAACACAGCTCAT
TGCAGATATCCAGGTTCTGAAACAAACCTCTAAACCATTAGCGAGTCCAGGAATCCAACCCAGTC
ACCAGAGCCAATTCTAGGGCAGTCGATTGATTATGAGAGACTCAATTATCAATATGACACA
CTTACCTCCACGGACATTCTATCTCAGAACATTGATGCCATGCTCAAGAGAGAAAGAAGGAAGAG
AGAACATTGCAAGCCTCTGTTGCACGGATTGGCGCCAGTGCTGATGTTGTTGATGTC
ACACCTGATGGTCATTGTGCCTTGCTGGAACCTCGCGTACTTGGTGGGAGCTTGAGATGCC
TGGCCTTGAAAGATTGTTCCGTACGATATCACAAAGGTTCTCAAGCAGATGAATCTTCACTAT
GATTCTGAGTTCCACTTGAGTTGAAGATTGTTGGCACAGATGGAACAGAACGCCATGGATCGT
ATCAAGAGCCCTACCATTGAACACCATTGGAGGAG

INTRON 2G/2H (SEQ ID NO:155)

GTATGTTTGAGATCCACATAATCTTCTACCCCTGTCTCATTCTAATGCTCTCAATACACAATT
ATATAGCCTTGAGCTTCAGATGATTACGGACAGGCATTACAGTATACATGTAATATGGTTTCT
GCTATTGCAAAATTGTCCTATCTGTTGAGATCAGTCATGGCGGTGACACCTAG

DOMAIN 2H (SEQ ID NO:159)

GTCACGATCACAGTGAACGTCACGATGGATTTTCAAGGAAGGAAGTCGGTCCCTGTCCCTGGATG
AAGCCAATGACCTAAAAATGCACGTACAAGCTGCAGAACATGATCAGGGTCCCAATGGATATGAAT
CAATAGCCGGTTACCATGGCTATCCATTCCCTGCGCTGAACATGGTAAGAACAGACTACGCATGCT
GTGTCCACGGAATGCCTGTATTCCACATTGGCACAGACTTCATACAATCCAGTTGAGAGAGCTC
TCAAAGAACATGGTTCTCATTGGGTCTGCCATACTGGGACTGGAC

Fig. 11aPrimary structure of the KLH2 proteinDOMAIN B

GLPYWDWTMPMSHLPELATSETYLDPPVTGETKNNPFHHAQVAFENGVTSRNPDAKLFMKPTYGDHT
 YLFDSMIYAFEQEDFCDFEVQYELTHNAIHAWVGSEKYSMSSLHYTAFDPIFYLHHSNVDRLWAI
 WQALQIIRRGSYKAHCASSQEREPLKPFAFSSPLNNNEKTYHNSVPTNVYDVGVLHYRYDDLQFG
 GMTMSELEEEYIHKTQHDRTFAGFFLSYIGTSASVDIFINREGHDKYKVGSFVVLGGSKEMKGFD
 RMYKYEITEALKTLNAVDDGFSITVEITDVDGSPPSADLI PPPAII FDVVR

DOMAIN C

ADAKDFGHSRKIRKAVDSLTVEEQTSLRAMADLQDDKTSGGFQQIAAFHGEPKWCPSPAEKKFA
 CCVHGMAVFPHWHRLLTQGENALRKHGFTGGLPYWDWTRPMSALPHVADPTYNDVSSEEDNP
 WYHGHIDSVGHDTRAVRDDLYQSPGFHYTDIAKQVLLAFEQDDFCDFEVQFEIAHNFIHALVGG
 NEPYSMSSLRYTTYDPIFFLHRSNTDRLWAIWQALQKYRGKPYNTANCAIASMRKPLQPFGQLDSVI
 NPDDETREHSVPFRVFDYKNNFDYEYESLAFNGLSIAQLDRELQRRKSHDRVFAGFLLHEIQSAL
 VKFYVCKHNVSDCDHYAGEFYILGDEAEMPWRYDRVYKYEITQQLHDLHVGDNFFLKYEAFDLN
 GGSLGGSIFSQPSVIFEPA

DOMAIN D

GSHQADEYREAVTSASHIRKNIRDLSERGEIESIRSAFLQIQKEGIYENIAKFHGKPGLEHDGHPV
 ACCVHGMPFPWHRLYVLQVENALLERGSAVAVPYWDWTEKADSLPSLINDATYFNSRSQTFDPN
 PFFRGHIAFENAVTSRDPQPELWDNKDFYENVMLALEQDNFCDFEIQLELIHNALHSRLGGRAKYS
 LSSLDYTAFDPVFFLHHANVDRIWAIWQDLQRYRKPYNEADCAVNEMRKPLQPFGNNPELNSDSMT
 LKHNLQDSDFYQNRFRYQYDNLQFNHSIQKLDQTIQARKQHDRVFAGFILHNIGTSAVVDIYIC
 VEQGGEQNCKTKAGSFTILGGETEMPFHFDRLYKFDITSALHKLGVPLDGHGFDIKVDVRAVNGSH
 LDQHILNEPSLLFVPGERKNIYY

DOMAIN E

DGLSQHNLVRKEVSSLTTLEKHFLRKALKNMQADDSPDGYQAIASFHALPPLCPSPSAHRHACCL
 HGMATFPQWHRLYTVQFEDSLKRHGSIVGLPYWDWLKPQSALPDLVTQETYEHLSHKTFPNPFLK
 ANIEFEGEVGTTTERDVDAEHLFAKGNLVYNNWFCNQALYALEQENYCDFEIQFEILHNGIHSWGG
 SKTHSIGHLHYASYDPLFYIHHSQTDRIWAIWQALQEHRGLSGKEAHCALEQMKDPLKPFSGSPY
 NLNKRTQEFSKPEDTFDYHRGYEDSLEFGMSVSSLHNYIKQQQEADRVFAGFLLKGFGQSASV
 SFDICRPDQSCQEAGYFSVLGGSSEMPWFDRLYKYDITKTLKDMKLRYDDTFTIKVHIKDIAGAE
 LDSDLIPTPSVLLEEGK

DOMAIN F

HGINVRHGRNRIRMELSELTERDLASLKSAMRSLQADDGVNGYQAIASFHGLPASCHDDEGHEIA
 CCIHGMPVFPHWRRLYTLQMDMALLSHGSAVAIPYWDWTKPISKLPDLFTSPEYYDPWRDAVVNNP
 FAKGYIKSEDAYTVRDPQDILYHLQDETGTSVLLDQTLALEQTDFCDFEVQFEVHNIAHYLVGG
 RQVYALSSQHYASYDPAFFIHHFSVDKIWAVWQALQKKRKRPYHKADCALNMMTKPMRPFAHDFNH
 NGFTKMHAVPNLFDQDLFYTDNLEIAGMNVNQLEAEINRRKSQTRVFAGFLLHGIGRSADVRF
 WICKTADDCHASGMIFILGGSKEMHWAYDRNFKYDITQALKAQSIHPEDVFDTDAPFFIKVEVHGV
 NKTALPSSAIAPPTIIYSAGE

Fig. 11b

DOMAIN G

DHIAGSGVRKDVTSLTASEIENLRHALQSVMDGGPNGFQAIAYHGSPPMCHMPDGRDVACCTHG
MASFPHWHLFVKQMEDALAAHGAHIGIPYWDWTSAFSHLPALVTDHEHNPFFHGHIAHRNVDTSR
SPRDMLFNDPEHGSESFFYRQVLLALEQTDFCQFEVQFEITHNAIHSWTGGHTPYGMSSLEYTAYD
PLFYLHHNSNTDRIWAIWQALQKYRGFQYNAAHCDIQVLKQPLKPFSESRNPNPVTRANSRAVDSFD
YERLNYQYDTLTFHGHISELDAMLQERKKEERTFAAFLLGFGASADVSFDVCTPDGHCAFAGTF
AVLGGELEMPWSFERLFTRYDITKVLKQMNLYHDSEFHFEKLIVGTDGTELPSDRIKSPTIEHHGG

DOMAIN H (SEQ ID NO:158)

GHDHSERHDGFFRKEVGSLSLDEANDLKNALYKLQNDQGPNGYESIAGYHGYPFLCPEHGEDQYAC
CVHGMPVFPWHRLHTIQFERALKEHGSHLGLPYWDW